Response to Comments from Editors and Referees

Information of previously submitted manuscript

- Manuscript number: egusphere-2023-1344
- Title: Evaluation of Calibration Performance of a Low-cost Particulate Matter Sensor Using Collocated and Distant NO_2
- Authors: Kabseok Ko, Ramesh R. Rao, and Seokheon Cho
- Status: Under Review

The authors would like to thank the referees for their careful reviews and insightful comments. We prepared responses for each comment from the editors and referees and revised our manuscript to reflect feedback.

Referee #1

The authors appreciate Referee #1's kind and valuable comments.

[Major Comments]

1. Section 3.1: 14 sensors were originally identified in this study, however only 5 were selected based on their months of valid measurements data. Of these 5, 2 were explicitly eliminated based on correlation analysis between the sensors' and their A and B units. Based on Figure 1 it seems like both PA-II 7 and 8 would be suitable for this study while PA-II 2, 3, 5, & 6 were not (Sensor 5 included in Figure 1 but not on line 188). Your final results will be more applicable if you are able to demonstrate improvements in more than 1 sensor, even if the study period is less than 2 years.

(Response)

We showed a series of results for PA-II 7 in our original manuscript. As the referee suggested, we evaluated PA-II 8's calibration performance under the following three cases:

Case 1: Calibration model is learned with the measurements collected from PA-II 8 in 2018 and calibration performance for the trained model is evaluated using data measured from PA-II 8 in 2019.

Case 2: This is similar to Case 1, except that the calibration model is trained with data measured from PA-II 7 in 2018.

Case 3: The measurement data from PA-II 8 with collocated NO_2 concentration in 2018 is used as a training dataset, while the data collected from PA-II 8 with either collocated NO_2 or distant NO_2 concentration in 2019 is used as a test dataset.

In Case 1, we evaluated the calibration model's performance with a test dataset consisting of measurement data from PA-II 8 in 2019. The calibration model is trained with data collected from the same PA-II 8 in 2018. Table R1 shows the calibration results of the PA-II 8 using a multiple linear regression (MLR) method under two different conditions: with and without NO_2. We selected the same feature vectors as defined in the original manuscript. We observed that NO_2 can enhance calibration performance because all MLR models using NO_2, except

combinations #10 and #11, yield lower errors and larger R^2 than those without NO_2. This observation aligns with the results shown in Table 3 of the original manuscript.

Additionally, compared to the calibration performance for PA-II 7 shown in Table 3 of the original manuscript, PA-II 8 shows slightly larger RMSE and MAE, but similar R^2.

Table R1. Calibration results of hourly $PM_{2.5}$ concentrations measured from the PA-II 8 in 2019 using MLR-based calibration model learned with training data collected from the PA-II 8 in 2018.

NO_2 not included				NO_2 included (i.e., collocated NO_2)			
Feature Vector	R^2	RMSE	MAE	Feature Vector	R^2	RMSE	MAE
1	0.731	4.559	3.468	10	0.741	4.468	3.381
2	0.749	4.397	3.299	11	0.742	4.459	3.375
3	0.760	4.307	3.223	12	0.783	4.087	2.982
4	0.763	4.277	3.191	13	0.785	4.072	2.966
5	0.759	4.311	3.219	14	0.788	4.042	2.951
6	0.762	4.281	3.185	15	0.788	4.040	2.950
7	0.767	4.242	3.143	16	0.785	4.071	2.970
8	0.768	4.227	3.121	17	0.785	4.071	2.967
9	0.770	4.214	3.128	18	0.791	4.015	2.915
				19	0.791	4.019	2.911
				20	0.792	4.006	2.895
				21	0.792	4.002	2.892

In Case 2, we evaluated the calibration model's performance using a training dataset collected from PA-II 7 in 2018, and a test dataset collected from PA-II 8 in 2019. Table R2 shows calibration results for PA-II 8 using the MLR method under two different conditions, such as with and without NO_2. As with the observation in Table R1, NO_2 is the key factor enhancing calibration performance. With the exceptions of #10 and 11, all MLR models using NO_2 yield lower errors and larger R^2 than those without NO_2. It is important to compare this result with that shown in Table 3 of the original manuscript, as we used different test datasets. It could be expected that the much worse performance for all feature combinations listed in Table R2 is achieved than for every corresponding feature vector in Table 3 of original manuscript, since the calibration model considered in Table R2 is tested with the data measured from the PA-II 8, whereas it is trained with the measurement data collected from the PA-II 7. R^2 values of all feature vectors in Table 10 are similar to those for each corresponding feature vector in Table 5. Unlike R^2, we observe larger RMSE and MAE when we populate the training dataset with measurements from PA-II 8 rather than PA-II 7. The maximum differences of RMSE and MAE for each feature vector in Tables 10 and 5 are 0.177 $\mu g/m^3$ and 0.196 $\mu g/m^3$, respectively.

The results shown in Tables R1 and R2 support our conclusion that reliable and consistent PA-II units, which contain two PMS 5003 sensors with high correlation, demonstrate similar calibration performance. This implies that a proposed calibration method can be applied to reliable and consistent PA-II units generally.

Table R2. Calibration results of hourly $PM_{2.5}$ concentrations measured from PA-II 8 in 2019 using MLR-based calibration model learned with training data collected from PA-II 7 in 2018.

NO_2 not included				NO_2 included (i.e., collocated NO_2)			
Feature Vector	R^2	RMSE	MAE	Feature Vector	R^2	RMSE	MAE
1	0.737	4.638	3.546	10	0.747	4.549	3.458
2	0.757	4.459	3.364	11	0.748	4.538	3.446
3	0.763	4.400	3.322	12	0.788	4.162	3.054
4	0.765	4.383	3.293	13	0.790	4.145	3.031
5	0.765	4.388	3.301	14	0.794	4.104	3.003
6	0.766	4.373	3.275	15	0.795	4.097	3.000
7	0.772	4.323	3.222	16	0.789	4.151	3.048
8	0.772	4.318	3.208	17	0.789	4.158	3.050
9	0.774	4.301	3.208	18	0.796	4.089	2.985
				19	0.795	4.100	2.984
				20	0.795	4.095	2.974
				21	0.796	4.090	2.970

Lastly, in Case 3, we evaluated the effect of collocated and distant NO_2 on the PA-II 8 unit's calibration performance. Table R3 shows the results of MLR-based calibration model for the PA-II 8 when it is verified with the test data considering either collocated or distant NO_2. As we explained in the original manuscript, we considered two monitoring sites measuring NO_2 near the Rubidoux site. One site (ID 06-065-8005) had NO_2 measurements that were much more highly correlated with the Rubidoux site than those from the other site (ID 06-071-00247). We refer to the NO_2 concentrations measured from these two sites as "distant NO_2." Three columns, describing the values of R^2, RMSE, and MAE, of collocated NO_2 in Table R3 are exactly the same as those of NO_2 included (i.e., collocated NO_2) in Table R1. In the case of site 06-065-8005 with high correlation to the Rubidoux site, the consideration of the distant NO_2 facilitates improvement of the calibration performance, since all MLR-based calibration models using distant NO_2, except combinations #10 and 11, produce lower errors and higher R^2 than those without NO_2. This result is similar to when we consider the collocated NO_2. However, we observe that adding distant NO_2 to the test dataset, which is not highly correlated to the NO_2 measurement from the reference site, deteriorates the calibration performance. This

is likely because all combinations from #10 to 21 yield lower R^2 and greater errors than all combinations excluding NO_2, as shown in Table R1. This result is the same as our observation of the PA-II 7 unit's calibration results in Table 5 of the original manuscript.

Hence, these results we draw from Table R3 support the same conclusions we drew from Tables R1 and R2. Reliable and consistent PA-II units achieve similar calibration performance, and our proposed calibration model can be applied to these units generally.

Table R3. Calibration results of hourly $PM_{2.5}$ concentrations taken from PA-II 8 in 2019 using an MLR-based calibration model learned with data collected from PA-II 8 in 2018 (Site ID indicates the monitoring sites for distant NO_2).

	MLR						
Site ID	Feature Vector	Collocated NO_2			Distant NO_2		
		R^2	RMSE	MAE	R^2	RMSE	MAE
6-065-8005	10	0.741	4.468	3.381	0.742	4.458	3.371
	11	0.742	4.459	3.375	0.744	4.442	3.359
	12	0.783	4.087	2.982	0.783	4.089	2.976
	13	0.785	4.072	2.966	0.786	4.066	2.951
	14	0.788	4.042	2.951	0.789	4.031	2.927
	15	0.788	4.040	2.950	0.789	4.033	2.930
	16	0.785	4.071	2.970	0.785	4.075	2.966
	17	0.785	4.071	2.967	0.785	4.076	2.960
	18	0.791	4.015	2.915	0.790	4.022	2.911
	19	0.791	4.019	2.911	0.790	4.026	2.908
	20	0.792	4.006	2.895	0.793	3.998	2.877
	21	0.792	4.002	2.892	0.793	3.995	2.875
6-071-0027	10	0.741	4.468	3.381	0.716	4.681	3.600
	11	0.742	4.459	3.375	0.716	4.680	3.591
	12	0.783	4.087	2.982	0.684	4.937	3.887
	13	0.785	4.072	2.966	0.684	4.937	3.864
	14	0.788	4.042	2.951	0.680	4.965	3.850
	15	0.788	4.040	2.950	0.672	5.030	3.914
	16	0.785	4.071	2.970	0.693	4.870	3.816
	17	0.785	4.071	2.967	0.706	4.764	3.704
	18	0.791	4.015	2.915	0.710	4.733	3.676
	19	0.791	4.019	2.911	0.713	4.705	3.646
	20	0.792	4.006	2.895	0.713	4.709	3.643
	21	0.792	4.002	2.892	0.699	4.818	3.756

We added the above study on Cases 1 to 3 as a new manuscript subsection, as follows:

3. 3 Applicability of Other PA-II Units

We evaluated PA-II 8's calibration performance under the following three cases:

Case 1: Calibration model is learned with the measurements collected from PA-II 8 in 2018 and calibration performance for the trained model is evaluated using data measured from PA-II 8 in 2019.

Case 2: This is similar to Case 1, except that the calibration model is trained with data measured from PA-II 7 in 2018.

Case 3: The measurement data from PA-II 8 with collocated NO_2 concentration in 2018 is used as a training dataset, while the data collected from PA-II 8 with either collocated NO_2 or distant NO_2 concentration in 2019 is used as a test dataset.

In Case 1, we evaluated the calibration model's performance with a test dataset consisting of measurement data from the PA-II 8 in 2019. The calibration model is trained with data collected from the same PA-II 8 in 2018. Table 9 shows the calibration results of the PA-II 8 using an MLR method under two different cases: with and without NO2. We selected the same feature vectors as defined in Table 4. We observed that NO2 can enhance calibration performance because all MLR models using NO2, except combinations #10 and #11, yield lower errors and larger R2 than those without NO2. This observation aligns with the results shown in Table 5. Additionally, compared to the calibration performance for PA-II 7 shown in Table 5, PA-II 8 shows slightly larger RMSE and MAE, but similar R2.

Table 9. Calibration results of hourly $PM_{2.5}$ concentrations measured from the PA-II 8 in 2019 using an MLR-based calibration model learned with training data collected from PA-II 8 in 2018.

NO_2 not included				NO_2 included (i.e., collocated NO_2)			
Feature Vector	R^2	RMSE	MAE	Feature Vector	R^2	RMSE	MAE
1	0.731	4.559	3.468	10	0.741	4.468	3.381
2	0.749	4.397	3.299	11	0.742	4.459	3.375
3	0.760	4.307	3.223	12	0.783	4.087	2.982
4	0.763	4.277	3.191	13	0.785	4.072	2.966
5	0.759	4.311	3.219	14	0.788	4.042	2.951
6	0.762	4.281	3.185	15	0.788	4.040	2.950
7	0.767	4.242	3.143	16	0.785	4.071	2.970
8	0.768	4.227	3.121	17	0.785	4.071	2.967
9	0.770	4.214	3.128	18	0.791	4.015	2.915

				19	0.791	4.019	2.911
				20	0.792	4.006	2.895
				21	0.792	4.002	2.892

In Case 2, we evaluated the calibration model's performance using a training dataset collected from PA-II 7 in 2018, and a test dataset collected from PA-II 8 in 2019. Table 10 shows calibration results for PA-II 8 using the MLR method under two different conditions, such as with and without NO_2. As with the observation in Table 9, NO_2 is the key factor enhancing calibration performance. With the exceptions of #10 and #11, all MLR models using NO_2 yield lower errors and larger R^2 than those without NO_2. It is important to compare this result with that shown in Table 5, as we used different test datasets. It could be expected that the much worse performance for all feature combinations listed in Table 10 is achieved than for every corresponding feature vector in Table 5, since the calibration model considered in Table 9 is tested with the data measured from the PA-II 8, whereas it is trained with the measurement data collected from the PA-II 7. R^2 values of all feature vectors in Table 10 are similar to those for each corresponding feature vector in Table 5. Unlike R^2, we observe larger RMSE and MAE when we populate the training dataset with measurements from PA-II 8 rather than PA-II 7. The maximum differences of RMSE and MAE for each feature vector in Tables 10 and 5 are 0.177 $\mu g/m^3$ and 0.196 $\mu g/m^3$, respectively.

The results shown in Tables 9 and 10 support our conclusion that reliable and consistent PA-II units, which contain two PMS 5003 sensors with high correlation, demonstrate similar calibration performance. This implies that a proposed calibration method can be applied to reliable and consistent PA-II units generally.

Table 10. Calibration results of hourly $PM_{2.5}$ concentrations measured from the PA-II 8 in 2019 using MLR-based calibration model learned with training data collected from the PA-II 7 in 2018.

NO_2 not included				NO_2 included (i.e., collocated NO_2)			
Feature Vector	R^2	RMSE	MAE	Feature Vector	R^2	RMSE	MAE
1	0.737	4.638	3.546	10	0.747	4.549	3.458
2	0.757	4.459	3.364	11	0.748	4.538	3.446
3	0.763	4.400	3.322	12	0.788	4.162	3.054
4	0.765	4.383	3.293	13	0.790	4.145	3.031
5	0.765	4.388	3.301	14	0.794	4.104	3.003
6	0.766	4.373	3.275	15	0.795	4.097	3.000
7	0.772	4.323	3.222	16	0.789	4.151	3.048
8	0.772	4.318	3.208	17	0.789	4.158	3.050
9	0.774	4.301	3.208	18	0.796	4.089	2.985

				19	0.795	4.100	2.984
				20	0.795	4.095	2.974
				21	0.796	4.090	2.970

Lastly, in Case 3, we evaluated the effect of collocated and distant NO_2 on the PA-II 8 unit's calibration performance. Table 11 shows the results of MLR-based calibration model for the PA-II 8 when it is verified with the test data considering either collocated or distant NO_2. As we explained in Section 3.2, we considered two monitoring sites measuring NO_2 near the Rubidoux site. One site (ID 06-065-8005) had NO_2 measurements that were much more highly correlated with the Rubidoux site than those from the other site (ID 06-071-00247). We refer to the NO_2 concentrations measured from these two sites as "distant NO_2." Three columns, describing the values of R^2, RMSE, and MAE, of collocated NO_2 in Table 11 are exactly the same as those of NO_2 included (i.e., collocated NO_2) in Table 9. In the case of site 06-065-8005 with high correlation to the Rubidoux site, the consideration of the distant NO_2 facilitates improvement of the calibration performance, since all MLR-based calibration models using distant NO_2, except combinations #10 and 11, produce lower errors and higher R^2 than those without NO_2. This result is similar to when we consider the collocated NO_2. However, we observe that adding distant NO_2 to the test dataset, which is not highly correlated to the NO_2 measurement from the reference site, deteriorates the calibration performance. This is likely because all combinations from #10 to 21 yield lower R^2 and greater errors than all combinations excluding NO_2, as shown in Table 9. This result is the same as our observation of the PA-II 7 unit's calibration results in Table 8.

Hence, these results we draw from Table 11 support the same conclusions we drew from Tables 9 and 10. Reliable and consistent PA-II units achieve similar calibration performance, and our proposed calibration model can be applied to these units generally.

Table 11. Calibration results of hourly $PM_{2.5}$ concentrations measured from PA-II 8 in 2019 using an MLR-based calibration model learned with training data collected from the PA-II 8 in 2018 (Site ID indicates the monitoring sites for distant NO_2).

	MLR						
Site ID	Feature Vector	Collocated NO_2			Distant NO_2		
		R^2	RMSE	MAE	R^2	RMSE	MAE
0 6 -	10	0.741	4.468	3.381	0.742	4.458	3.371
	11	0.742	4.459	3.375	0.744	4.442	3.359
	12	0.783	4.087	2.982	0.783	4.089	2.976
	13	0.785	4.072	2.966	0.786	4.066	2.951

065-8005	14	0.788	4.042	2.951	0.789	4.031	2.927
	15	0.788	4.040	2.950	0.789	4.033	2.930
	16	0.785	4.071	2.970	0.785	4.075	2.966
	17	0.785	4.071	2.967	0.785	4.076	2.960
	18	0.791	4.015	2.915	0.790	4.022	2.911
	19	0.791	4.019	2.911	0.790	4.026	2.908
	20	0.792	4.006	2.895	0.793	3.998	2.877
	21	0.792	4.002	2.892	0.793	3.995	2.875
06-0711-0027	10	0.741	4.468	3.381	0.716	4.681	3.600
	11	0.742	4.459	3.375	0.716	4.680	3.591
	12	0.783	4.087	2.982	0.684	4.937	3.887
	13	0.785	4.072	2.966	0.684	4.937	3.864
	14	0.788	4.042	2.951	0.680	4.965	3.850
	15	0.788	4.040	2.950	0.672	5.030	3.914
	16	0.785	4.071	2.970	0.693	4.870	3.816
	17	0.785	4.071	2.967	0.706	4.764	3.704
	18	0.791	4.015	2.915	0.710	4.733	3.676
	19	0.791	4.019	2.911	0.713	4.705	3.646
	20	0.792	4.006	2.895	0.713	4.709	3.643
	21	0.792	4.002	2.892	0.699	4.818	3.756

We added the following subsection on the effect of various training periods:

3.4 Effect of Training Period

We evaluated the effect of the training period on calibration performances. We consider four different training periods (i.e., 3, 6, 9, and 12 months), and each training set is constructed as follows: The training sets all end at the close of 2018. Their start points are set in reverse order based on training periods. For example, for 3 months, the training set is from Oct. 2018 to Dec. 2018. Table S4 shows PA-II 7's calibration results using the MLR method for all four training periods. The 3-month training period had the worst performance. The 6- and 9-month training periods generated better performances than the 12-month training period. From a viewpoint of using NO_2, NO_2 can improve calibration performance in all four cases, compared to using only temperature and relative humidity. As the length of the training period increases, calibration performance improves.

Table S4 Effect of training period length on calibration performance using MLR-based calibration model

	3			6			9			12		
Feature Vector	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
1	0.738	4.447	3.193	0.750	4.344	3.204	0.729	4.524	3.436	0.731	4.513	3.418
2	0.746	4.382	3.132	0.775	4.125	2.925	0.763	4.231	3.110	0.755	4.305	3.194
3	0.752	4.329	3.094	0.776	4.111	2.971	0.761	4.253	3.163	0.760	4.263	3.165

4	0.757	4.286	3.051	0.779	4.083	2.930	0.762	4.237	3.136	0.763	4.232	3.132
5	0.743	4.411	3.171	0.779	4.090	2.906	0.767	4.193	3.080	0.763	4.234	3.129
6	0.745	4.393	3.151	0.782	4.061	2.869	0.769	4.182	3.056	0.765	4.211	3.100
7	0.749	4.355	3.127	0.787	4.012	2.828	0.779	4.091	2.964	0.772	4.154	3.043
8	0.748	4.364	3.073	0.786	4.020	2.823	0.779	4.087	2.950	0.772	4.151	3.023
9	0.742	4.420	3.058	0.783	4.051	2.798	0.778	4.101	2.940	0.771	4.161	3.012
10	0.735	4.477	3.181	0.762	4.243	3.100	0.736	4.470	3.383	0.741	4.424	3.329
11	0.721	4.592	3.208	0.762	4.246	3.101	0.736	4.468	3.382	0.741	4.423	3.326
12	0.774	4.133	2.882	0.806	3.832	2.651	0.794	3.946	2.805	0.789	3.997	2.871
13	0.762	4.244	2.880	0.806	3.832	2.639	0.794	3.946	2.806	0.789	3.993	2.857
14	0.764	4.226	2.874	0.811	3.782	2.579	0.798	3.908	2.787	0.792	3.962	2.843
15	0.762	4.241	2.888	0.811	3.775	2.572	0.799	3.900	2.783	0.793	3.954	2.838
16	0.769	4.178	2.949	0.805	3.842	2.646	0.794	3.950	2.805	0.790	3.986	2.866
17	0.769	4.175	2.944	0.806	3.833	2.631	0.793	3.960	2.804	0.789	3.990	2.863
18	0.775	4.126	2.919	0.807	3.821	2.663	0.805	3.840	2.693	0.798	3.912	2.793
19	0.770	4.168	2.900	0.806	3.831	2.668	0.805	3.841	2.693	0.796	3.925	2.790
20	0.770	4.170	2.904	0.803	3.858	2.690	0.805	3.835	2.687	0.797	3.920	2.782
21	0.765	4.218	2.936	0.803	3.860	2.692	0.805	3.837	2.690	0.797	3.913	2.777

2. Line 298: What is the reasoning behind this 1:1 data split, specifically using the whole year of 2018 to train the models and apply to 2019. This implies that in practice you have to wait a whole year before collecting valid/corrected data with this method which hinders the use of low-cost sensors. And assuming minimal sensor drift from 2018 to 2019 and similar environmental conditions.

(Response)

The 1:1 data split reflects seasonal patterns in $PM_{2.5}$ and other environmental parameters, such as temperature and relative humidity. To more accurately gauge the relationship between a PA-II unit and regulatory measurements across seasons, we used whole-year data for training. To support our efforts, we studied the training period's effect on calibration performance. As shown in Table S4, training with a shorter period like 3 months yields lower RMSE and MAE than training with 6, 9, or 12-month periods. Hence, it is necessary to train calibration models with data collected over a long enough period to fully account for seasonality and provide reliable performance.

Over time, the degradation of electrical components or dust accumulation can cause drift in low-cost PM sensors. Figure R1 shows $PM_{2.5}$ concentrations obtained from both PA-II 7 and 8 units whose internal PMS 5003 sensors have high correlation with each other. Both PA-II units render similar $PM_{2.5}$ concentrations over time, which makes it challenging to verify the amount of drift experienced by each unit. Therefore, we assume that each PA-II unit has different and minimal drift. Under this assumption, when we compare the performance of the two calibration models, which are trained with distinct datasets from the PA-II 7 and PA-II 8 units in 2018, respectively, and verified with the same test dataset collected from PA-II 8 in 2019, minimal drift has a minor effect on calibration performance, since both units demonstrate similar calibration performance through RMSE and MAE. This comparison was described and explained in Tables 9 and 10.

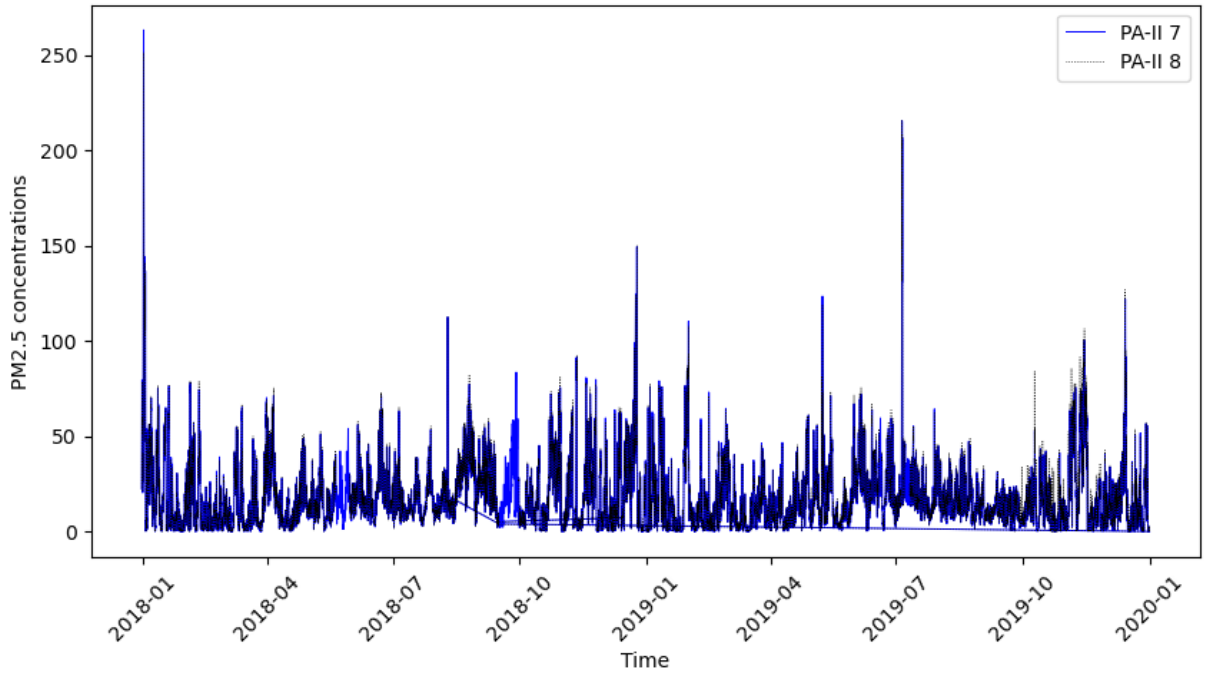

Figure R1. Results of $PM_{2.5}$ concentrations from both PA-II 7 and 8 units

[Minor Comments]

1. Figure 1: Please include info about PA sensors A and B in the caption as you did on line 193.

(Response)

We updated the caption for Figure 1 as follows:

Correlation among all PMS 5003 sensors of the selected units PA-II 2, 3, 5, 6, 7, and 8. The left and right of each number on the x-axis represent PMS A and B sensors for its corresponding PA-II unit, respectively.

2. Figure 2: Include a 1:1 line for comparison.

(Response)

We added a 1:1 line to Figure 1.

3. Figure 3: Ensure x-axes are the same for the PM2.5 graph and temperature+RH graph. Figure sizes could be increased to improve readability.

(Response)

We modified the x-axes in the two subfigures to reflect this suggestion.

4. Line 36: Please clarify that FRM and FEM are US EPA designations and may not be applicable to every county.

(Response)

We updated the sentence on line 36 as follows:

The monitoring stations use instruments based on Federal Reference Methods (FRMs) or Federal Equivalent Methods (FEMs), which promote high precision and accuracy. The U.S. Environmental Protection Agency (EPA) approves both FRMs and FEMs as official designations for measuring ambient concentrations. Furthermore, the U.S. EPA carries out various cooperative programs, including those on ambient monitoring methods and technologies, with many other countries in the world.

5. Line 61: "good a correlation" Please correct to "a good correlation".

(Response)

We modified our text as recommended.

6. Line 74: More discussion needed on how NO2 contributes to PM2.5 formation.

(Response)

We updated the sentence on line 74 as follows:

In addition to these direct factors, we examine the impact of the precursor gas NO_2, acting as a source of $PM_{2.5}$ emissions, on calibration performance in low-cost $PM_{2.5}$ sensors. In general, $PM_{2.5}$ arises by secondary formation from a chemical reaction between precursor gases,

such as NO_2, in the atmosphere some distance downwind from the original emission source (Hodan et al., 2004).

{Reference}: Hodan, W.H. and Barnard, W.R.: Evaluating the Contribution of PM2.5 Precursor Gases and Re-entrained Road Emissions to Mobile Source PM2.5 Particulate Matter Emissions, MACTEC Federal Programs, 2004.

7. Line 127: Typo for US EPA

(Response)

We modified it as recommended.

8. Line 131: What is the purpose of the 2-minute vs 80 sec interval?

(Response)

We updated the sentence on line 131 as follows:

In the first step, when we calculate 1-hour averages of $PM_{2.5}$ measurements generated with 2 min (or 80 sec) intervals, we remove the 1-hour average if the number of $PM_{2.5}$ measurements is less than 27 (or 40). We considered two different measurement intervals for a PA-II unit because its old interval had been 80 sec until May 30, 2019. Its current interval is 2 min.

9. Line 178: Please clarify the difference between the FRM instrument and the BAM instrument. Does the FRM only report daily values?

(Response)

We updated the sentence on line 178 as follows:

The monitoring site we considered has an FRM instrument and a BAM-1020 instrument with the parameter of 88502. These instruments produce daily and hourly $PM_{2.5}$ measurement data, respectively. Since we measure the PA-II units at intervals much shorter than a full day, it is much more reasonable to compare the $PM_{2.5}$ measurement of PA-II units with that of a

BAM-1020 instrument with a shorter measurement interval, rather than an FRM instrument for evaluating the accurate calibration performance of PA-II units. However, we face the limitation that a BAM-1020 instrument can be classified as a non-FEM-compliant device. Therefore, our approach for analyzing PA-II units to appropriately resolve these issues is as follows: we compared the BAM-1020 instrument's readings with daily $PM_{2.5}$ concentrations collected from an FRM instrument to ensure the BAM-1020 provides an acceptable level of performance as an FRM instrument, which is enough to assess the calibration performance of PA-II units. According to this affirmative observation, the BAM-1020 instrument can be used to evaluate the calibration performance of low-cost $PM_{2.5}$ sensors by comparing its readings with hourly $PM_{2.5}$ measurement data of PA-II units.

Also, we updated the sentence on line 203 as follows:

These data suggest that a BAM-1020 instrument using non-FEM methods compares well to the statistics achieved with the FRM method.

10. Line 206 + 236: You list 6 significant figures/3 decimal points for several of the PA-II sensors, yet these sensors are not that accurate. As per the manufacturer +/-10 ug/m3 for 0-100 ug/m3 and +/-10% for 100-500 ug/m3. Please correct.

(Response)

We deleted Lines 206 and 235. Instead, we added the following footnote in Table 3 describing summary statistics of daily and hourly $PM_{2.5}$ concentrations from a FRM instrument, a BAM-1020 instrument, and a PA-II unit:

[Footnote] A PMS 5003 sensor that collects $PM_{2.5}$ concentrations from within a PA-II unit exhibits a maximum consistency error of +/-10 $\mu g/m^3$ at 0-100 $\mu g/m^3$ and +/-10% at 100-500 $\mu g/m^3$. The sensor reports $PM_{2.5}$ concentrations as integer values on a per-second basis. A PA-II unit generates readings of its own $PM_{2.5}$ concentrations by averaging its 1-second $PM_{2.5}$ concentrations over 80 (or 120) seconds. In this study, daily (hourly) $PM_{2.5}$ concentrations are calculated by averaging $PM_{2.5}$ concentrations rendered by a PA-II unit over 24 hours (1 hour), and thus can be represented with a decimal number. In other words, the presence of decimal

numbers in daily and hourly $PM_{2.5}$ concentrations reported by the PA-II 7 unit does not indicate precise concentration measurements.

11. Line 219: How are you defining the r correlation of 0.928 as "good"?

(Response)

We updated the sentence on line 209 as follows:

In this study, we examined the root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), and Pearson correlation coefficient, *r,* between daily PM2.5 data from the FRM instrument and that from the PA-II units. In the cases of the RMSE, MSE, and MAE, the lower its value is, the better the performance or the lower the difference in measurement data between the FRM instrument and the PA-II units. The Pearson correlation coefficient is a metric measuring a linear correlation between two variables. It is a number between -1 and 1 that measures the strength and direction of their relationship. As the coefficient approaches an absolute value of 1, the values of measurement data from the FRM instrument and the PA-II units becomes more similar.

We updated the sentence on line 219 as follows:

These results show that the PA-II unit has a good correlation (r) with the FRM instrument for the two-year period of interest, since its value is very close to 1.

12. Line 220: You say performance of FRM and BAM did not correlate favorably, yet in line 203 you state that the non-FEM method compared well to FRM? Why do you conclude that the BAM is less favorably correlated to the FRM when its statistics are better than the PAs?

(Response)

We modified Line 220 as follows:

However, a comparison of metrics from the FRM instrument and the PA-II unit did not correlate as favorably.

13. Line 230: Please clarify why the FRM instrument was not used to evaluate hourly performance? Were hourly FRM measurements not available?

(Response)

We updated the sentence on line 233 as follows:

Next, we compared the PA-II unit's hourly $PM_{2.5}$ data with that of the BAM-1020 instrument over the course of the same two-year period. We did not consider the FRM instrument for exploring hourly $PM_{2.5}$ measurement data, since it only produces daily concentrations.

14. Line 272: The referenced article does not actually consider NO2 in their PM2.5 calibration. They only used PM2.5, Temperature, RH, CO, and wind speed in their models.

(Response)

The author of the article (Hua et al., 2021) claimed that $PM_{2.5}$ exhibits positive associations with NO_2, which indicates that NO_2 emissions make a large contribution to $PM_{2.5}$ pollution in the winter.

15. Line 293: "because month has a different slope..." Do you mean " because each month..."?

(Response)

We updated the sentence on line 293 as follows:

It is challenging to use the per-month linear fitting result to calibrate PA-II units because each month has a different slope and intercept defined for the linear fitting.

16. Lines 311 + 355: Can these lists be included as Tables rather than in-text to improve readability and when readers look at Tables 3-5.

(Response)

We added Tables for listing the selected feature vectors as a referee suggested.

17. Line 395: "Corresponding R2 values did not differ meaningfully" Based on what statistics, do you have a *p*-value?

(Response)

We updated the sentence on line 394 as follows:

For instance, the RMSE values from the best MLR and RF models were 3.912 $\mu g/m^3$ and 3.840 $\mu g/m^3$, respectively. Their corresponding R^2 values differ slightly, since their gap is only 0.008. Nonetheless, the MAE of 2.777 $\mu g/m^3$ achieved from the best MLR is lower than that achieved by the best RF, which is 2.831 $\mu g/m^3$.

18. Line 408: How are you defining moderate and high correlations?

(Response)

We updated the sentence on line 408 as follows:

The site 06-065-8005 had NO_2 measurements that are much more highly correlated with the Rubidoux site compared with those from the site 06-071-0027. This result can occur when the distance from the Rubidoux site to the site 06-065-8005 is shorter than it is to the site 06-071-0027.

19. Line 412: "We used NO2 for training a calibration model" Which NO2 data to train from, from Rubidoux? Please clarify.

(Response)

We rewrote Lines 410-413 as follows:

To evaluate the usefulness of distant NO_2 measurements on the calibration of a low-cost PM sensor, we used NO_2 data measured from monitoring sites near the PA-II 7 unit as a test dataset, rather than data from the collocated Rubidoux site. When we trained calibration models with the measurements from the PA-II 7 unit over 2018, we used highly accurate NO_2 concentrations

measured by FEM instruments at the Rubidoux site. Subsequently, to verify the trained calibration models, we utilized a separate test dataset featuring distant NO_2 measurements taken by FEM instruments at sites 06-065-8005 and 06-071-0027. We considered this scenario to evaluate our proposed calibration models, previously trained with collocated NO_2 concentrations and distant NO_2 concentrations, when collocated NO_2 measurements cannot be collected.

20. Line 430: "but not significantly" Based on what statistics, do you have a p-value?

(Response)

We updated the sentence on line 423 as follows:

All MLR methods using distant NO_2 data from site 06-071-0027 had a higher RMSE than the MLR algorithm was based on data that did not include NO_2 data from the collocated Rubidoux instrument, which had an RMSE of 4.513 $\mu g/m^3$ as shown in Table 5.

We updated the sentence on line 430 as follows:

In the case of RF models, the use of the distant NO_2 data from site 06-065-8005 increased RMSE compared to using collocated NO_2 data, but not significantly, since the maximum gap of RMSE values for all feature vectors considered was just 0.060 $\mu g/m^3$.

21. Line 447: Please re-word sentence as the point is unclear.

(Response)

We updated the sentence on line 447 as follows:

The factors, directly affecting the performance of a low-cost PM sensor, including temperature, relative humidity, and particle composition, have been scrutinized for their impact on sensors' performance enhancement.

22. Line 448: Please re-word to clarify that the inclusion of NO2 as an environmental factor in the calibration has potential to improve...

(Response)

We updated the sentence on line 448 as follows:

Additionally, this study investigated the potential of NO_2, a precursor gas that gives rise to $PM_{2.5}$ through atmospheric chemical reactions, to improve performance of the calibration model.

23. Section 2.2 Please include more information about the monitoring instrumentation used, especially the NO2 monitoring sites.

(Response)

We updated the sentence on line 110 as follows:

Monitoring ambient air quality for purposes of determining compliance with the U.S. National Ambient Air Quality Standards (NAAQSs) requires the use of either FRMs or FEMs. FRM and FEM instruments are accepted as methods for monitoring the NAAQS pollutants, including particulate matters (i.e., $PM_{2.5}$ and PM_{10}), NO_2, SO_2, O_3, and CO. Hourly measurements of $PM_{2.5}$, and other pollutants, such as NO_2, SO_2, O_3, and CO, obtained from FEM and non-FEM instruments can be downloaded via the EPA's application programming interface (https://aqs.epa.gov/data/api) (U.S. EPA, 2011).

{Reference}: U.S. EPA: Reference and Equivalent Method Applications: Guidelines for Applicants, Sep. 2011.

24. Section 3.2 + 3.3: At various points you include or drop units for your RMSE, MSE, MAE and r stats. Please be consistent. Shouldn't r (R2) be unitless? Please be consistent in using r vs R2.

(Response)

We modified these units throughout the paper, as recommended.

25. Section 3.6.3: Please check units of ug/m3 as you often have "ugm3" in this section.

(Response)

We modified these units throughout the paper, as recommended.

26. Equations 3, 4, & 5 could be included in the methods section rather than results.

(Response)

As recommended, we created a new subsection called Performance Evaluation Metrics, and moved the relevant paragraph to a new subsection.