Response for RC2: egusphere-2023-969 "Performance evaluation of MOMA – a remote network calibration technique for PM2.5 and PM10 sensors"

We thank the reviewer for the constructive and helpful comments. The reviewer's comments were addressed point-by-point. Our response is highlighted in red below and in yellow in the manuscript.

General comments:

This paper uses two calibration approaches (one using an approach of calibrating at a monthly interval, one using a test system to calibrate given certain criteria are met) using low cost sensors and nearby "proxy" sites. The approaches build off previous work and seem to provide good performance, although comparisons to other methods are lacking. The article is generally well written with a few sections noted below needing some clarification and attention to units.

Specific comments:

For readability, consider changing the names of the various AQY sensors (e.g. AQY BD-1146) to something more related to their deployment (e.g. AQY RIVR 1) or use a simple number scheme (e.g. AQY 1) for use within the paper. A table with the original names could be provided in the supporting information.

We have re-labelled the co-located AQYs and added this information to Table 1.

Section 2.3: Can the authors comment on what percentage of the data included fog and was thus discarded from the datasets? Additionally, do other processes that might impact visibility such as wildfires pose a risk for removing data from the calibration periods? Is the main issue with the fog from issues with hygroscopicity (or more generally from high humidity) or with visibility itself? Looking at Figure S1 for the month of November 2021 it seems like a small portion of the dataset was excluded (maybe 10% or less). I recommend that this percentage (either total or broken down by month) be mentioned in the article.

We have added information about data completeness and percentage of fog (<1% for each site) to Table 1. The issue with fog arises from high humidity and hygroscopicity which lead to overestimates of the PM2.5 and PM10 concentrations from the sensor but not the co-located reference.

Section 2.3: Could the authors comment on how the drift calibration approach deals with periods dominated by local sources? Are there any checks that are made to determine if the reason for a calibration alarm is due to a local source that would not be picked up by a proxy? For the purposes of this study the collocated reference monitors can be used to verify, but for future deployments is there a protocol?

We used a 3-day running averaging window to calculate the three statistical tests used to determine if a sensor has drifted. This averaging period was selected to smooth short-term local effects while retaining diurnal and regional variations. At the same time, we use a 5-day window to trigger drift and a calibration, thus a threshold needs to be exceeded for continuous 5-days before a sensor a gets calibrated (L176ff).

Section 2.3: What is meant by a "suitable seven-day calibration window" for the monthly calibration approach? Other than removing fog and ensuring data completeness, are there other metrics for

choosing the most appropriate window out of the two week period of consideration? Additionally, is this window chosen with the help of the next month's data or is made completely independently of the next month's data?

We have provided further details regarding the selection criteria of the seven-day calibration window as well as how new gains and offsets are applied (L168 – L171). The calibration was performed independently of the next month's data to test the performance of MOMA as a framework for real-time sensor calibrations.

Section 2.4: What are the criteria of classification for roadways used in the analysis? For instance, is the distance of the site from a motorway mean to the nearest highway or does any road at all count?

We used the distance to the nearest highway. Primary roads and highways (L187) were included when counting the road length within a 1 km buffer.

Section 3.2: How was the most similar land use proxy site determined? I do not see any data related to this (e.g., the metrics discussed in section 2.4 related to roadways).

We have added a sentence how we selected the proxy with most similar land use (L190).

On line 207 the authors state that the nearest proxy was generally more useful than the one with most similar land use (which is unsurprising). Why did the authors choose to use the nearest proxy for all sites rather than the ones which performed the best and have a mix of nearest and similar land use? If it is for simplicity, I would suggest mentioning that.

While it is possible to use the proxy that performed 'best' at sites with a co-located reference there is no measure to confirm this at sites with no co-located reference or prior data. Thus, we applied a general rule that can be used for non-co-located sensors for future deployments (L242).

Section 3.3.1: I would suggest including metrics on the performance statistics discussed on line 224 as "good" can mean different things to different readers. This also applies to line 279.

We have modified the statements (L267 ff)

Section 3.4: Figure 10 needs an explanation in the caption or the legend for the difference between the step changes and the continuous curves.

We have added a description to the caption of Figure 10 (Figure 6 in the revised version).

Section 4: The article could benefit from a table summarizing the performance of the calibration approaches so the reader does not need to use multiple figures to determine the efficiency of each approach (this table could be in the results section and summarized briefly in the conclusions). Additionally, a comparison to other methods of calibration should be discussed. The strengths and weaknesses of the monthly and drift approaches should be discussed relative to these other options.

We have replaced the scatterplots with a summary table to display the performance metrics used throughout the paper (Table 4). A comparison to other methods and a brief summary of the strengths and weaknesses of the monthly and drift approaches has been added to the conclusions (L417 – 435).

Technical corrections:

Consider adding units to Figure 1 a and b for latitude and longitude (e.g. [° N]).

Latitude and longitude coordinates given in decimals do not need units as the decimal representation itself implies the units.

Line 41, consider changing “with particle type or properties changes over time” to “with particle type or their properties may change over time”.

We revised the sentence (L46).

Line 59, consider writing out “Los Angeles” for the first time. It is written out on line 74 currently.

Done.

Line 70, I believe the non-regulatory air monitors discussed in this line are the AQY systems discussed in section 2.2. Consider clarifying this in section 2.1.

Done.

Line 104, consider changing “drive” to “driven”. Done.

Line 167, consider replacing the comma between PM25 and PM10 with “and”.

This statement is no longer included as we moved the figure to the SI.

For Figure 2 consider lining up the various sites vertically and leaving gaps for missing data so that they can be easily compared between the graphs. Not critical but would help the reader compare.

We modified the figure, this figure is now in the SI (SI Figure 2).

Line 191, consider adding in “respectively” after R2. We revised this sentence.

Line 215, consider adding units to the distance columns of Table 1. Done.

Line 264, Figure 5. Some of the equations on the individual panels are cut off. Same for Figure 7. We removed the scatterplots and display the results in a table instead (Table 4).

Line 256, consider changing to “as often is the case”. Done.

Line 275, Figure 6 caption. Add in the metric being plotted to the description of panel a (i.e. MAE).

Done.

Line 314, Figure 9. The a0 panel should include units of µg m-3.

Done.

Line 319, add in a closing parenthesis at the end of the line. Done.