Comments on Cross et al. (2017) from N. Zimmerman, R. Subramanian, A. Presto and A. Robinson, CMU

General Comments

This paper discusses using HDMR to calibrate the low-cost sensors used in the Aerodyne ARISense air quality monitor. While the results seem promising, it is difficult to assess the performance of the model, because the training data appear to have been included as part of the model performance assessment. This would bias the model performance and makes it difficult to compare the results with other studies that evaluate model performance using independent datasets.

Additionally, we believe the paper would benefit from more discussion on building and interpreting the HDMR model. Questions such as what was the maximum order used, what variables were significant, and any physical interpretation of any significant variables are either missing or underdeveloped. The paper would also benefit from some additional metrics of model performance beyond correlation plots.

Another question to address is how the training data are chosen. From Figure 6, it appears that only periods where there were pollutant concentrations were elevated were chosen to build the model. How could this calibration approach be generalized for others? If the training data set was carefully constructed vs. randomly selected then is it feasible to assume that the model training window could be condensed to 1 week, as the other reviewers suggest?

As a full disclosure, we are also in the process of submitting a manuscript on a different type of calibration model for low-cost electrochemical sensors. We welcome and encourage feedback from Aerodyne on our manuscript in kind to help the community collectively improve sensor performance.

Specific Comments

Page 5: Line 26-27: Can you be more specific? What is your definition of "acceptable accuracy" –the paper would benefit greatly from some quantitative performance metrics.

Page 6, Line 11-12: What is the statistical analysis done to decide which variables are significant? Something like AIC/BIC? ANOVA? T-test?

Page 6, Line 13-14: I am not sure I fully understand the HDMR. Can the orthogonal basis functions be written in closed form (parametric?) I think a couple extra sentences here introducing the model are warranted.

Page 6: Line 20-23: What is the spanned range? For others building their own co-location windows, what were the critical criteria to determine the optimal co-location period? Was 35% arbitrarily chosen or was the calibration window tuned and if so, what was learned during tuning? Some discussion of diminishing returns vs. training window would be helpful to others implementing these methods.

Page 6 Line 12-18: This is another paragraph where I think some quantitative performance criteria would be useful. When comparing the performance of HDMR calibrations to manufacturer corrections or corrections by other papers, it's not clear what the terms 'reasonably good correlation' or 'relatively small' mean.

Page 7 Line 26-27: It seems like a lot of interesting work was done in the lab, but none of these results are provided. I'd be interested to see more details here. Can this be included in supplemental?

Page 8, Line 14: What was the environmental variability spanned? And how was the 35% subset chosen? This is a follow up to the previous comment.

Page 8, Line 20: This seems problematic, was the performance of the model tested on a data set in which 35% was used for training? Ideally the model should be tested on completely blind test data (i.e., the remaining 65%). If this is what you did, it should be made clearer. If this is not what you did, you should provide performance metrics for the pure testing data since this approach is the only way to truly test the model performance.