

**Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements, Cross et al., submitted to AMTD 5 May, 2017.**

We thank the reviewers for their careful reading of the manuscript and many helpful comments. Our responses to specific comments can be found below. The reviewers' comments are in italics, our responses are in plain font, and changes made to the manuscript are in quotation marks and indented. Page and line numbers refer to the original document. We have made significant changes to some of the figures and have included new performance metrics for the HDMR model. However, none of these changes affect the conclusions in the manuscript.

**Response to Reviewer RC1: Anonymous Referee #1**

*This is a rigorous attempt to illustrate the challenges and utility of deploying 'low cost' air quality sensors in a community. This is a field of growing interest – likely to become more crowded – with important implications across most spheres of atmospheric research. A particularly strength of this work is stressed by the authors in warranting caution in interpretation of data from these types of sensors.*

*Specific comments: P1, Line 14: perhaps this is better phrased as '...address environmental justice issues related to air quality.' P1, Line 22: 'Protecting the air environment...' isn't really one of the most important PH challenges. Rather, it is protecting populations from degraded air quality exposure that is important. P1, Line 29: to presume the authors mean US dollars? P8, Line 18: extra word 'in'*

We have made the suggested changes to wording.

P1, line 14: "...address environmental justice issues related to air quality."

P1, line 22: "Protecting populations from exposure to poor air quality is..."

P1, line 29: "...thousands of US dollars."

P8, line 18: deleted "in"

*General Comments: One issue that is not discussed is the potential for lot variability in sensor performance within a single manufacturer. Whilst the authors provide adequate detail on which make/model has been chosen (P2, L33-40), do these EC sensors exhibit differences within a manufacturer production lot? Or are there differences across different lots?*

This is an important question, and one that lies at the heart of the low-cost sensor calibration challenge. Based on our empirical experience, we have observed significant (up to a factor of 2.5x) differences in sensitivity for batches of otherwise identical sensors. In these cases, the sensors themselves have the same age (relative to manufacturer production and out-of-package timescales). This evidence supports the need to build sensor-specific (and in the context of multi-pollutant measurement systems, system-specific) calibrations as opposed to a general network-wide calibration. Based on our limited observations, the manufacturing process for these electrochemical sensors is not yet fully reproducible.

We have revised the end of Section 2.1 (page 3) to address this issue:

"This paper presents results for the four electrochemical sensors in a single ARISense system. Note that nominally identical electrochemical sensors can have widely different sensitivities and

exhibit variable environmental interference effects. As a result, the specific calibration models described in this paper cannot be broadly applied to all ARISense systems. Until the reproducibility of electrochemical sensor manufacturing improves, system-specific HDMR models will need to be developed for each individual ARISense system to maintain sensor quantification metrics.”

*The paper begins with a discussion on environmental justice (abstract), and includes very specific references to asthma rates in the sampled community (P4, lines 1-4). Context is, of course, important, but these facts seem out of place in this manuscript which is mainly a focus on the technical details of using and interpreting EC sensors.*

We have removed some of the details about asthma rates. The end of the first paragraph on page 4 now reads:

“The original DAQSS deployment and initial ARISense proof-of-concept efforts were motivated by the need to assess the viability of lower-cost AQ sensor systems in communities suffering from environmental health knowledge gaps, such as the unexplained doubling of the adult asthma rate in North Dorchester between 2001 and 2010 (Backus, 2012).”

*It was surprising to see a reported temperature range of 5-45 degrees in a northern US city, but the authors later state that this was internal box temperature to assess electrode function under actual operating conditions. When comparing these data, was a correction to ambient temp and RH taken into account (e.g. temp/RH measured by a nearby met station)? For example, if ambient temp and RH were 25 deg and 50%, but the internal temp where 35 deg and 15% because of strong sunlight, one would expect a significant effect, given the apparent sensitivity.*

Due to the sensitivity of the electrochemical sensors, it is more important to consider measurements of temperature and RH at the sampling interface of the sensors rather than ambient conditions, as these internal sensor-specific conditions will most closely correspond to the interference signal observed by the working and auxiliary electrodes. However, comparison of sensor system temperature and RH inside the flow-cell with ambient temperature and RH showed that even under conditions of direct sunlight, the sensor system internal conditions remained within 10-15% of the ambient values throughout the co-location sampling period. We have expanded our explanation of the temperature and RH measurements in the 4<sup>th</sup> paragraph of section 2.1 as follows:

“The manifold includes an embedded RH/T sensor positioned adjacent to the electrochemical cells which is used to model the temperature and relative humidity-derived interference effects on the raw sensor response. Given the active flow of the gas sampling inlet and minimal residence time (~1s) of the sample air within the manifold, the RH and T measurements recorded by the ARISense system closely track changes in ambient RH and T conditions. Over the co-location period described here, measurements inside the flow manifold were within 10-15% of the ambient values even under conditions of direct sunlight.”

*Why was there no data included or discussed for particulate matter or CO2?*

We explained at the end of the third paragraph on page 3 that the particulate matter measurements will be assessed in a future manuscript. Due to the size detection limit for the OPC, and the anticipated size distribution of near-field and accumulation mode aerosol particles in most urban environments – the utility and reliability of low-cost OPCs for PM<sub>2.5</sub> measurements remains highly uncertain. As such, significant analysis effort is required to reconcile OPC metrics and this is the subject of ongoing work in

our laboratory. Since this paper focuses on calibration of electrochemical sensors, we chose not to include the CO<sub>2</sub> data. We have added a sentence at the end of the second paragraph on page 3 explaining this:

“Note that the CO<sub>2</sub> measurements are not discussed in this paper which focuses on the electrochemical sensors, but will be addressed in a future manuscript.”

*In a number of cases, the authors refer to this sensor package as a ‘low-cost’ replacement for measuring air quality, which could play a key role towards empowering environmental justice (P9, Line 34). The authors are correct in asserting that lower cost sensors likely have a role in improving granularity in air quality monitoring networks, especially in locations with disproportionate air quality burdens, like the relatively low income communities in which this study takes place. But the idea of ‘low cost’ is a fairly subjective statement that seems meant to broaden the appeal of these products to communities in need. The development of low or lower cost sensor units with an eye towards reducing injustices is a noble and important direction for air quality scientists, but it might provide value to compare this instrument against the few other existing low/lower cost sensing units that are found in the literature – both in terms of sensor performance and relative cost.*

The reviewer makes an excellent point that “low-cost” is a subjective term. We have changed the designation for ARISense throughout the paper from “low-cost” to “lower-cost,” and have added the following sentences in the introduction to explain the rough cost tiers for air quality monitoring systems. It is difficult to provide the exact cost of other Tier 2 systems (integrated multi-pollutant measurement packages) since this information is often proprietary. We have also included a table summarizing recent published performance data for lower cost AQ systems that utilize Alphasense electrochemical sensors in Table 4.

“Air quality monitoring systems can be roughly divided into three cost tiers, 1) high cost/high accuracy systems costing tens to hundreds of thousands of US dollars, such as those used at regulatory monitoring stations, 2) lower cost systems costing a few to ten thousand US dollars, such as the ARISense system or the recently developed Real-time Affordable Multi-Pollutant (RAMP) package developed by Carnegie Mellon University and Sensever (Zimmerman et al., 2017), and 3) low cost systems (costing tens to hundreds of US dollars) designed for the consumer market that typically only measure a single pollutant and generally suffer from poor quality data (EPA, 2017). The goal of second tier systems is to provide data quality approaching Tier 1 at a fraction of the cost.”

*The largest issue seems to be in interpretation and setup of the HDMMR model to adjust sensor data to real values. Specifically, the authors state that the model can ‘capture the intricate interdependencies of the variables. . .’ in order to correct the data and provide guidance to researchers which variables are most impactful (P6, Line 1). These statements presume that the researchers enter in all possible variables that are likely to play a role in sensor performance. Given the relative few number of variables measured, and presumably computed, how can a researcher have confidence that they are accounting for all – or at least most – of the likely variants that may affect their results? The concern here is that there may be other plausible covariates that affect sensor performance. For example, one might imagine a measure of CO<sub>2</sub> by NDIR could be affected by water vapor (which is imputed by this sensor package), but also by other ambient IR-absorbing components (that are not measured)?*

To re-phrase the reviewer’s question – they are asking about the challenge of unknown unknowns – if there is a specific interference vector that is not explicitly measured (and modeled) with the system, how can we be confident that the results from the integrated system will remain robust in the real-world

presence of this interference vector? To first order, we are confident that the model can handle the full extent of environmental interference vectors encountered in the Dorchester micro-environment based on the results of the test data which leverages the long-term co-location of the sensor with reference measurements. By combining on-board measurements of multiple pollutant species (via raw sensor outputs – both WE and Aux electrode signals) and environmental conditions (P, T, RH) and allowing the ambient variability of these species and conditions to train the model, the results suggest that the primary factors impacting sensor performance are captured by the model for this set of electrochemical sensors (although clearly, the Ox-B421 sensor is underperforming relative to the others). It is certainly possible that the NDIR measurement of CO<sub>2</sub> could be impacted by cross-sensitivity to species not measured by the other sensors in the system, but modeling and validating the NDIR sensor response is beyond the scope of the current manuscript and will be examined in a subsequent work.

*The authors also note (P 6, Line 8-9) that in the first step of modeling, a user can choose how many variables are selected to interact with one another. How does one quantitatively make this determination?*

This comment overlaps with comments from Reviewers 2 and 3, requesting more information on how the model was optimized. We have added the following paragraph on page 7 and to describe the development of the HDMR model for the NO sensor:

“An example of how the HDMR model is developed for the NO-B4 sensor is provided in the Supplemental Material. The left column of Table S2 lists all available input parameters and the other columns denote which parameters were included in the input matrix for each model run. The bottom rows list the RMSE, MAE, and MBE for each model run for both the training data (model generation) and test data (model evaluation).”

We have added the following text, table and figure to the Supplemental Material:

“Table S2 shows a subset of the input matrices for training the NO-B4 sensor output to the NO reference measurements. Six model versions are shown (labelled v.8-13), and the resultant RMSE, MAE, and MBE (in ppb) are listed at the bottom of the table for both the training set and the test data (with test metrics shown in curly brackets). The optimal model run (v.8) is indicated with shading. The table shows that while the model with the most diverse set of inputs (v.12) resulted in the lowest RMSE, MAE, and MBE values for the training data, its RMSE and MAE were worse compared to v.8 when applied to the ambient test data. It should be noted that *ExploreHD* also performs statistical F-tests to further refine which inputs and input pairs are considered in the HDMR model training, and to determine a suitable degree for polynomial basis functions for each component function.

The poorer performance of model v.12, trained with the full set of inputs available can be explained by increased overfitting related to the additional degrees of freedom from the increased number of input pairs. The F-tests performed by *ExploreHD* during model generation are aimed at mitigating issues with overfitting, but only consider each input independently. Thus, this automated input selection is not perfect, especially for cases like electrochemical sensor quantification, where there is significant correlation between certain inputs in the training data. The approach used here of testing a range of input sets, effectively serves as a manual supplement to the automated input selection performed by *ExploreHD*.

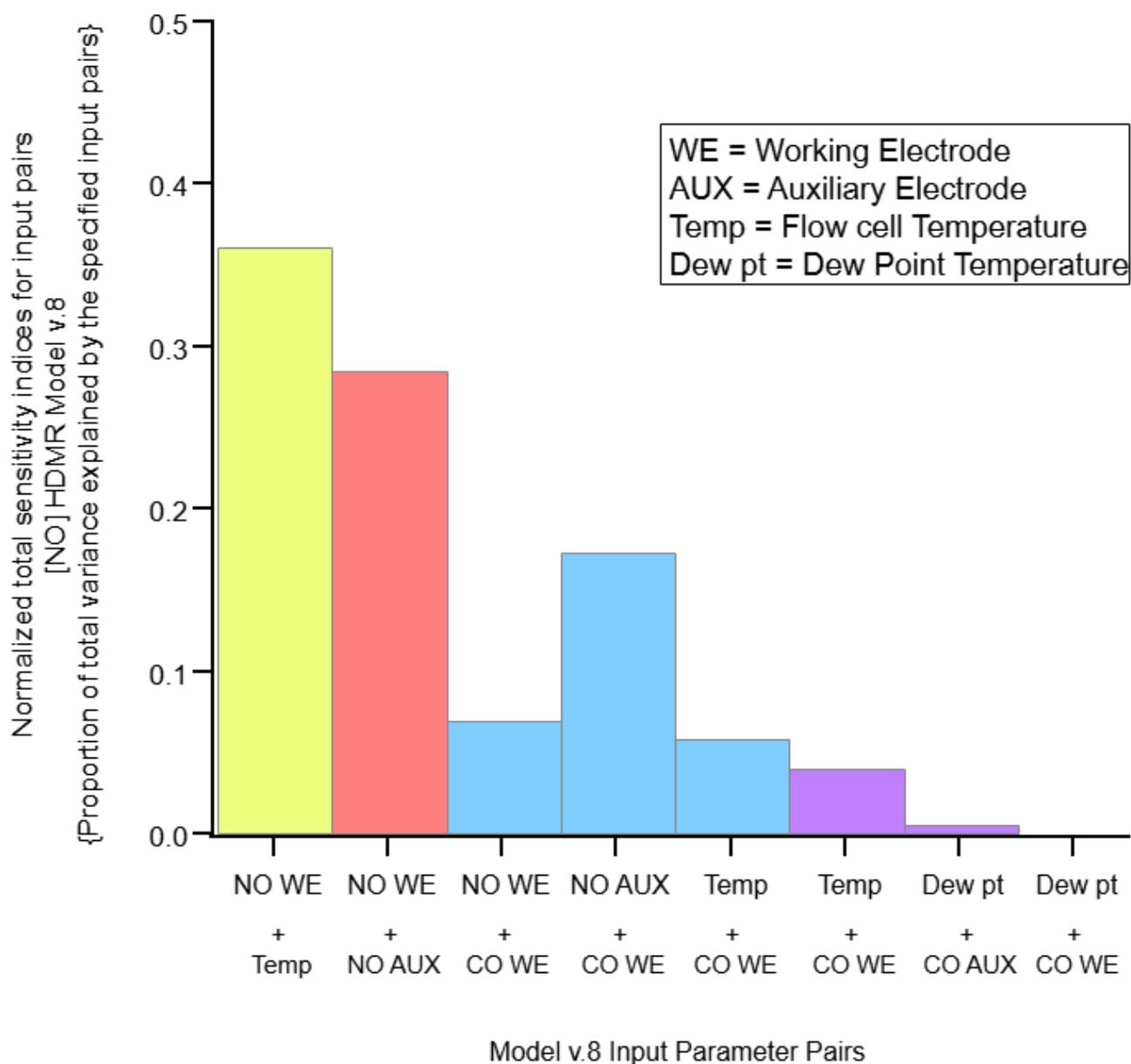
From Table S2 it is also seen that models excluding key inputs (*e.g.* v.10) exhibit poorer performance on test (and training) data. The input selection used in model v.8 exhibits a reasonable trade-off between the issues of exclusion of important inputs and overfitting (as the

performance on training and test datasets were comparable in this case). Through future work, it may be possible to refine or replace the F-test-based input selection algorithm used by *ExploreHD* so that overfitting might be addressed in a more automated fashion for training datasets exhibiting high correlation between certain inputs.

Figure S1 shows key input pairs in the NO HDMR v.8 model. The figure plots normalized total sensitivity indices for the input pairs. These sensitivity indices quantify the proportion of variance that can be explained by each input pair, considering both structural and correlative components. These metrics are the result of a structural and correlative sensitivity analysis (SCSA) performed by *ExploreHD*, which is described in [Li et al. “Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs”, *J. Phys. Chem. A.* **114**, 2010, 6022]. In addition to calculating a total sensitivity index for each input / input pair, this analysis decomposes the total sensitivity into structural contributions reflecting the underlying system model, and correlative contributions reflecting covariation between inputs in the dataset being considered. Decomposition of sensitivities in this manner provides the opportunity for additional insights into the role of each input / input-pair.”

**Table S2. Set of HDMR models for NO sensor. The optimal model (v.8) is indicated with shading.**

Model v.	8	9	10	11	12	13
CO AUX	x				x	x
CO WE	x			x	x	x
NO AUX	x	x			x	x
NO WE	x	x	x	x	x	x
NO2 AUX					x	x
NO2 WE					x	x
Ox AUX					x	
Ox WE					x	
Dew point	x	x	x	x	x	x
Temperature	x	x	x	x	x	x
CO2 (voltage)					x	x
RMSE (ppb)	3.38	5.09	6.75	4.88	2.58	2.81
{ test }	{ 4.52 }	{ 5.86 }	{ 7.16 }	{ 5.53 }	{ 9.19 }	{ 6.41 }
MAE (ppb)	2.40	3.29	4.23	3.05	1.63	1.76
{ test }	{ 2.83 }	{ 3.90 }	{ 4.94 }	{ 3.27 }	{ 4.07 }	{ 3.27 }
MBE (ppb)	0.02	0.32	-0.55	0.15	-0.01	-0.05
{ test }	{ 0.97 }	{ 1.80 }	{ 2.08 }	{ 1.08 }	{ 0.30 }	{ 0.87 }



“**Figure S1.** Normalized total sensitivity indices of each significant (contribution > 0.1%) input pair in NO Model v.8. Of the possible combinations, the NO-WE/Temp, NO-WE/NO-AUX, and NO-AUX/CO-WE explain more than 80% of the sensor-system variance trained against the corresponding reference [NO] measurement.”

*It is very difficult to discern useful results from Figure 3. Further, we must presume that these data have been validated by the investigators. If so, it is surprising to see spikes of ozone exceeding 1000ppb with some regularity in this location, as observed by the reference monitor.*

We agree with the reviewer that Figure 3 is not very useful. We have therefore removed it and included a 72-hour segment of the differential voltages in what was formerly Figure 5 (now Figure 3). The reference data for O<sub>3</sub> plotted in the original Figure 3d was incorrect (displaying reference CO data instead of O<sub>3</sub> data in the AMTD version of the paper). We apologize for this error.

*Figure 5 is a fairly useful illustrative figure that clearly identifies sensor limitations. But it is troubling to see divergence between the EC sensor and the reference sensor in periods of relative stability in temperature. This seems to need further explanation – how does your data compare for this specific time series after it has been modeled?*

Temperature is not the only variable driving the divergence between the EC and reference sensors, hence the need for a complex, multi-dimensional model to explain the variance in the raw sensor relative to reference. We have revised Figure 5 (now Figure 3) to include the model output and we have revised the text on page 7 (now page 8) as follows:

“Figure 3 shows the time-series for a ~ 72-hour period for the relative humidity, dew point temperature (panel a, solid and dashed lines, respectively), temperature (grey shaded area), and raw differential sensor output (dashed line), reference measurement (thick red dashed line) and model output (thin solid line) for the four electrochemical sensors (panels b-e). The raw differential sensor output is displayed as a voltage ( $\Delta mV$ ) which is linearly proportional to the difference in current generated within the electrochemical cell at each electrode (working and auxiliary). The correlation plots between the raw EC-sensor output and the reference measurements are shown in Figs. 4a to d, with each data point colored by flow-cell temperature. The intercept, slope and  $r^2$  for the linear regression indicated with a red line in Fig. 4 are listed in Table 1.”

with additional discussion on page 10:

“Closer examination of the model output for 72-hours of the test data in Fig. 3 gives additional clues for improving the model. In Fig. 3d at ~18:00 on 11/2/2016, the model NO<sub>2</sub> exceeds the reference NO<sub>2</sub> by a factor of ~2 during a period of rapidly decreasing temperature and increasing RH. This underscores that the rate of change of input parameters may be important in the model, in addition to the absolute values. Fig. 3e also suggests that the HDMR model for O<sub>x</sub> struggles during times of rapidly changing temperature, particularly when the O<sub>3</sub> concentration is low (< 3 ppb). Future development of HDMR models to support ARISense quantification will include derivatives of key variables as inputs.”

*The authors included a number of variables to consider in adjusting or training the model, but specifically excluded sensor age, noting that the sensors were approximately 6-7 months old at the end of the study and, therefore, should have limited effect on model performance. Firstly, wouldn't it be more appropriate to compare sensor age to manufacturing date, rather than when a package is opened?*

Based on communications with the manufacturer of the sensors, sensor aging is directly related to the loss of electrolyte (7M H<sub>2</sub>SO<sub>4</sub>) from a given EC cell, which is primarily driven by exposure to extremes in RH (< 15% or > 80%). Under these conditions the electrolyte will either evaporate (<15%) from the cell or absorb significant H<sub>2</sub>O (> 80%), overflowing the cell. Upon completion of a batch of sensors, the

manufacturer ships the cells in self-contained sealed containers at 25C and 60% RH, following equilibration in their laboratory. Therefore, we interpret the onset of sensor-aging (sensor t<sub>0</sub>) as the date at which this seal is broken for a given individual sensor.

*And second, it is unsatisfying to ignore sensor age as a relevant variable, given the relatively short lifetimes of these sensors. For example, the NDIR lamp and electrode has a lifetime of 2000-6000 hours (according to the manufacturer), depending on lamp light time and the presence of heavy contaminating pollution. This is 80- 250 days, which is not much longer than the study length presented here, and suggests that long term drive is something that should not be ignored.*

While we agree that modeling the decay (i.e., aging) of electrochemical sensors is extremely important, such an exploration is beyond the scope of the analysis presented here. The manufacturer quotes the following operating lifetimes (degradation of signal to 50%): 36 months for CO, and 24 months for NO, NO<sub>2</sub> and O<sub>x</sub>. These timescales are longer than the 4.5 month deployment (6.5 mo. out of package) pertaining to the current work, and we therefore do not expect significant (>5%) sensor degradation due to aging. A detailed assessment of sensor performance over 18-24 months of continuous ambient operation is ongoing and it is with this subsequent dataset that time will be considered as an input to the model in an effort to track and correct for degradation in performance over sensor lifetime. The NDIR sensor is not discussed in this manuscript. We have revised the last paragraph of section 2 (page 8) to read:

“The data presented in this paper were recorded over a 4.5-month sampling interval (July 7, 2016- November 23, 2016). All four electrochemical sensors used in this study were first removed from their packaging on May 9, 2016. That means that from out-of-package, the sensors had aged ~6.5 months by November 23. The manufacturer quoted lifetime for degradation of the signal to 50% is 36 months for the CO sensor and 24 months for the NO, NO<sub>2</sub> and O<sub>x</sub> sensors. Given that these lifetimes are significantly longer than the deployment time scale analysed here, we did not include a time-dependent sensitivity term in the input matrix of our HDMR model runs. The results presented here therefore assume that the sensitivity of each of the electrochemical sensors did not appreciably drift over the 4.5 month deployment. In subsequent studies we will analyze sensor response over longer deployment timescales (18 to 24 mo.) to investigate the importance of including a time-based parameter to track and correct for drift in sensor response with time.”

## **Response to Reviewer RC2: Anonymous Referee #2**

*General comments: This paper is timely in describing how to improve the performance of a set of Alphasense electrochemical sensors, which are being widely incorporated into many emerging multipollutant air quality sensor technologies. The paper goes into great depth in exploring causes of sensor measurement artifacts and demonstrates an approach to improve the data quality. However, this paper will have a limited impact if several important issues are not addressed. A recommendation of major changes is suggested, focusing upon these areas of improvement:*

- 1. How are authors defining “good enough” for sensor data quality? They indicate a goal of having credible data and “acceptable accuracy” (line 27), but need to clarify what they consider to be their target (accuracy, measurement range, etc.) and for what purpose.*

We thank the reviewer for this important question and have revised the paper to include the performance metrics of root mean square error (RMSE), mean absolute error (MAE) and mean bias error (MBE). We have added the following paragraph on page 7 and the following table to the Supplemental Material.

“The metrics used to evaluate the model are the slope and intercept of a linear least squares regression of the model output with the reference measurements, the coefficient of determination of the linear fit ( $r^2$ ), the root mean square error (RMSE), the mean absolute error (MAE), and the mean bias error (MBE). The equations for these metrics are given in Table S1 and model-to-measurement results are summarized in Tables 2, 3, and 4.”

**Table S1. Metrics used for comparing EC sensor model output ( $y_i$ ) to reference measurements ( $x_i$ ).**

Statistic	Abbrev.	Formula	Description
Coefficient of determination	$r^2$	$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	$f_i$ is the value of the linear least squares fit at $x_i$ . Ratio of explained variation to total variation. For linear least squares regression, $r$ is equal to Pearson's correlation coefficient.
Root mean square error	RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$	Standard deviation of difference between model output and reference values. Measure of accuracy. Sensitive to outliers.
Mean absolute error	MAE	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - x_i $	Average of the absolute error. Disregards the direction of under- or over-prediction.
Mean bias error	MBE	$MBE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)$	Average error. Indicates if model output values are biased high or low relative to reference values.

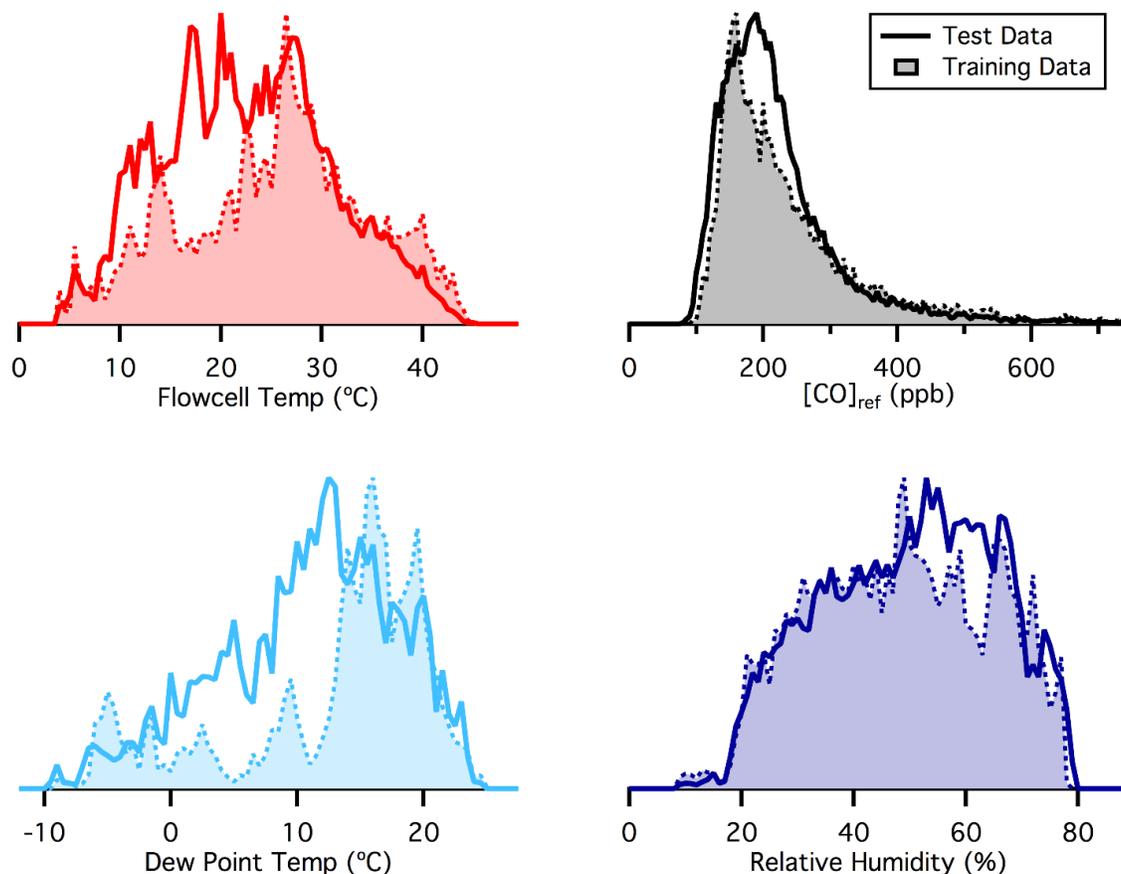
2. The authors note in their concluding sentence that “This compression of the training period is especially important. . .” Currently, they used 35% of a 4 month period of data to develop a complex model to improve the data. Why 35%? What is the performance if only 10% of the data were used? What if only the first week of data were used? Authors have sufficient data to explore the implications of different training periods that would provide important insight to researchers looking to employ sensors and develop study plans yielding reasonable data quality. It is recommended that authors go into substantially more depth to investigate the training period required.

The goal of this work is to show that with a sufficiently wide range of input parameters, we can successfully use the HDMR model to analyze electrochemical sensor raw output data. We did not systematically attempt to minimize the training period. Instead, we focused this initial modeling effort on evaluating whether the HDMR model could yield robust results when trained with a fairly conservative (intentionally comprehensive) set of input data (incl. range of gas concentrations, magnitude of environmental conditions and rates of change of environmental conditions). However, in our re-analysis of the test and training data points prompted by this question, we realized that we had unnecessarily eliminated test data by requiring that all four reference measurements be sampling ambient simultaneously (with none of the reference instruments in auto-calibration mode). Because the reference instruments all have different calibration schedules, this inadvertently decreased the set of test data. By treating each individual reference measurement separately, we were able to recover ~ 10% of the total test data and use that to evaluate the HDMR models. As a result, the training data now constitutes ~ 24-27%

of the total eligible co-location data points, leaving ~73-76% of the data available to test the HDMR model against ambient pollutant/condition variability.

We have also added a figure to the supplemental material to show the distribution of parameters used during the training and test periods. And we have added the following text on page 9 to clarify our goals:

“The training data for the HDMR model were chosen to provide comprehensive coverage of environmental variability spanning the July-November sampling interval. It was important to include (1) sensor responses to the range of gas concentrations encountered in ambient air (near-zero to high concentration transient spikes in pollution), (2) the range of temperatures and various rates-of-change in temperature, and (3) the range of measured water content of the sample air in the flow-cell. The goal was to include a wide enough range of training data to avoid extrapolation errors when applying the model to the test dataset (all ambient co-location data not included in the training dataset). Figure S3 shows the distributions of temperature, reference measurement, dew point temperature and relative humidity for the training data for the CO-B4 HDMR model, overlaid with corresponding distributions of the test data. We did not attempt to minimize the amount of ambient data used for training, or vary the timing of the training data with respect to the test data. Approximately 25% of the full time series was used to generate the model (Table 2 and indicated with grey bars in Fig. 5). The exact fraction of data used for training was slightly different for each sensor due to differing calibration schedules for the reference measurements (which automatically excludes sensor data from the training or test datasets). For each sensor, the set of inputs included in the input data matrix was optimized as described in Section 2.5 and the Supplemental Materials.”



**Figure S3.** Distributions of temperature, reference measurement, dew point temperature, and relative humidity for the training data for the CO HDMR model (dashed lines/shaded) and the test (solid lines) data. In the case of CO, the training data distributions were generated from 27% of the total available co-located interval with 7974 5-min average data points comprising the model training matrix and 21533 5-min average data points comprising the test data.

3. *Authors should investigate an aging effect – they indicate they will only explore this later, but should at minimum demonstrate whether there is any relationship with the number of “out of box” or “in use” days. In Jiao et al (2016, <https://doi.org/10.5194/amt-9-5281-2016>), aging was clearly demonstrated in a number of sensor types that incorporate Alphasense sensors.*

As noted in the response to Reviewer 1, the deployment time (4.5 months) was much shorter than the manufacturer quoted lifetimes (24 to 36 months) of the sensors. A detailed assessment of sensor performance over 18-24 months of continuous ambient operation is ongoing and it is with this subsequent dataset that sensor age will be considered as an input to the HDMR model formulation in an effort to track and correct for degradation in performance over sensor lifetime.

4. *How variable is the performance between identical sensors? How variable are the HDMR models from one RAMP to another?*

This work focused on results obtained from a single ARISense system, which is different from the Real-time Affordable Multi-Pollutant (RAMP) sensor package developed by Sensever and Carnegie Mellon

University for which multiple units were evaluated in Zimmerman et al., 2017. As mentioned in response to Reviewer 1, there is significant variability at the manufacturing level of Alphasense electrochemical sensors, which results in the need to build sensor-specific calibration models across the same type (e.g., CO-B4) of sensor. Moving forward we anticipate that unique HDMR models will be necessary for each ARISense system, due to the irreproducibility of the manufacturing process as it relates to each electrochemical sensor. The aim of the current work is to provide the first demonstration of the HDMR model trained on an ambient co-location dataset. Future laboratory-based efforts will focus on building HDMR models for multiple nodes from a compressed (~1 week) training interval during which pollutant concentrations and environmental conditions are systematically varied to sample the full range of conditions relevant to the field.

We have revised the end of Section 2.1 (page 3) to clarify this issue:

“This paper presents results for the four electrochemical sensors in a single ARISense system. Note that nominally identical electrochemical sensors can have widely different sensitivities and exhibit variable environmental interference effects. As a result, the specific calibration models described in this paper cannot be broadly applied to all ARISense systems. Until the reproducibility of electrochemical sensor manufacturing improves, system-specific HDMR models will need to be developed for each individual ARISense system to maintain sensor quantification metrics.”

*5. The HDMR analysis is fairly opaque – authors cite papers that describe the approach, but do not provide sufficient detail for this to be reproducible. It is recommended that authors provide more specific information on the HDMR analysis and resulting model in the supplemental information. Given some sensor applications involve real-time transmission and display of data to the public, does the HDMR approach support this or must it be performed post hoc?*

As discussed above in the response to Reviewer 1, we have added a description of how the HDMR model is developed in the main text with supporting tables in the Supplemental Material. Real-time implementation of the HDMR model is feasible by adding the sensor-specific and system-specific models (in closed form, algebraic expressions) to the backend database architecture to enable real-time reporting of calculated ppb values.

*Minor comments:*

- *Quality of the text on figures needs improvement – recommend not using red font text and ensuring clear, readable axes.*

The red font text within figures has been modified and the font for axis labels has been increased in size and bolded.

*Authors compare against DEP monitors – they should indicate what are the detection limits of the monitors and implications for their calibration. Since regulatory monitoring stations are employed to evaluate air quality relative to the NAAQS, detection limits can be an issue in low concentration areas (e.g., some CO monitors have ~300 ppb detection limits, which may be fine for the NAAQS at a ppm level but may be an issue for co-location and calibration of sensors to be used for low-ambient sampling).*

We have modified section 2.3 to include the limit of detection (LOD) and RMS noise for each of the reference monitors used in the current study.

“The reference measurements used in this study include ozone (O<sub>3</sub>, Teledyne Model T400 Photometric Ozone Analyzer, LOD <0.6 ppb; RMS < 0.3 ppb), carbon monoxide (CO, Teledyne Model 300EU Carbon Monoxide Analyzer, LOD < 20 ppb; RMS ≤ 10 ppb), and nitrogen oxides (NO, NO<sub>x</sub>, NO<sub>2</sub>, Teledyne Model T200 Nitrogen Oxide Analyzer, LOD = 0.4 ppb; RMS < 0.2 ppb).”

*Abstract has some awkward statements that could be improved, as well as providing more quantitative results. e.g., “live, work, breathe. . .” – breathing is something that happens at all locations. . .one would hope. Also what is meant by “stakeholders”? The public? Industry?*

We have deleted the word “breathe” in the first sentence of the abstract. We have replaced “stakeholders” with “public.”

*Did the authors ever characterize the response time of the sensors? (e.g., against high time-resolution instruments also made by Aerodyne). A brief statement on their utility for a mobile sampling approach and time base of the data would be helpful, as many low cost sensor systems are being employed in a mobile fashion.*

ARISense v1.0 was designed to serve as a stationary, fixed site AQ node. Given that the fastest response times we could access from the DEP monitoring station equipment in this work was 60s averages, the time-response of the sensors themselves did not lag the rates of change in pollutant concentrations that were characterized by the reference instrumentation at the site. Laboratory studies are underway that will examine the response time of the electrochemical sensors more carefully and evaluate the limits of reconciling pollutant gradients from mobile measurement platforms (bicycle and drone), recording sensor metrics at 1Hz. This work will be the subject of a forthcoming publication.

**Response to Reviewer SC1: N. Zimmerman, R. Subramanian, A. Presto and A. Robinson**

*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.*

We thank the reviewers for bringing this issue to our attention. This was a serious oversight on our part and we have now corrected it. We have now re-evaluated the model performance using the test data only (i.e., excluding the training data). The results are presented in Figures 3, 4 and 5 and the performance metrics are in Table 3. Excluding the training data from the test data decreased the  $r^2$  for each sensor by approximately 10%, but did not change the conclusions of the paper.

*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.*

As discussed in the response to Reviewer 1, we have expanded our discussion of building and interpreting the HDMR model in the main text and in the Supplemental Material.

*The paper would also benefit from some additional metrics of model performance beyond correlation plots.*

We agree with the reviewers that additional performance metrics besides the slope of the correlation plot and  $r^2$  are important to include. As noted in the response to Reviewer 1, we have included RMSE, MAE and MBE as additional performance metrics.

*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 noted above in the response to Reviewer 2, we chose training data that covered the full range of the parameters that would be encountered in the test data. We have added additional discussion in the main text and Figure S3 with the distributions of parameters measured during the training and test data periods.

*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.*

We look forward to commenting on your manuscript.

#### *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.*

Quantitative performance metrics are now included for the training and test datasets, listed in Tables 2 and 3 of the main text.

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

HDMMR uses F-test as an initial evaluation of the relative importance of individual input parameters to a given trained output vector. As noted above in the response to Reviewer 1, a more thorough description of our approach to HDMMR is now included in the Supplemental Materials.

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

Yes, the 2<sup>nd</sup> order polynomial fits (cubic as max) can be exported from the HDMMR log files into a closed form, algebraic equation which in turn can be embedded in the backend database architecture of the ARISense server to provide real-time concentration metrics through the online user-interface. These equations are also embedded in the firmware for each ARISense system so that concentration values can be logged to the local on-board USB drive if the system is run in off-line mode.

*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.*

As noted above in the response to Reviewer 2, we did not attempt to minimize or tune the training window. Our goal was to choose a set of training data that covered the full range of the input parameters that would be encountered during the test data.

*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.*

We have added a table to the main text (Table 4) summarizing results from three recent studies examining Alphasense electrochemical sensor-derived concentrations, tested against co-located reference measurements in ambient urban and suburban micro-environments.

*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?*

While we agree with the reviewer that, in many ways, the laboratory setting provides a controlled environment across which the sensor response to a matrix of conditions can be characterized, our laboratory work with the ARISense system is ongoing and will be the subject of a forthcoming manuscript. The aim of that work will be to demonstrate that compressed (~1 week) training datasets can be generated through systematic laboratory experiments and that the resultant HDMR models provide a robust approach to ambient pollutant measurements made with the ARISense system across a variety of relevant micro-environments.

*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.*

These questions are answered above.

*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.*

We agree with the reviewers that including the training data in the testing data was problematic and regret the error.

# Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements

Eben S. Cross<sup>1</sup>, Leah R. Williams<sup>1</sup>, David K. Lewis<sup>1,2</sup>, Gregory R. Magoon<sup>1</sup>, Timothy B. Onasch<sup>1</sup>, Michael L. Kaminsky<sup>3</sup>, Douglas R. Worsnop<sup>1</sup> and John T. Jayne<sup>1</sup>

<sup>1</sup>Center for Aerosol and Cloud Chemistry, Aerodyne Research, Inc., Billerica, MA 01821 USA

<sup>2</sup>Department of Chemistry, Connecticut College, New London, CT 06320 USA

<sup>3</sup>Massachusetts Institute of Technology, Cambridge, MA 02139 USA

Correspondence to: Eben Cross (escross@aerodyne.com)

**Abstract.** The environments in which we live, work, and play are subject to enormous variability in air pollutant concentrations. To adequately characterize air quality, measurements must be fast (real-time), scalable, and reliable (with known accuracy, precision, and stability over time). Lower-cost AQ sensor technologies offer new opportunities for fast and distributed measurements, but a persistent characterization gap remains when it comes to evaluating sensor performance under realistic environmental sampling conditions. This limits our ability to inform the public about pollution sources and inspire policy makers to address environmental justice issues related to air quality. In this paper, initial results obtained with a recently developed lower-cost air quality sensor system are reported. In this project, data were acquired with the ARISense integrated sensor package over a 4-month time interval during which the sensor system was co-located with a state-operated (Massachusetts, USA) air quality monitoring station equipped with reference instrumentation measuring the same pollutant species. This paper focuses on validating electrochemical sensor measurements of CO, NO, NO<sub>2</sub>, and O<sub>3</sub> at an urban neighborhood site with pollutant concentration ranges (5-min averages,  $\pm 1\sigma$ ): [CO] = 23 ± 116 ppb; {spanning 84-1706 ppb}, [NO] = 6.1 ± 11.5 ppb; {spanning 0-209 ppb}, [NO<sub>2</sub>] = 11.7 ± 8.3 ppb; {spanning 0-71 ppb}, and [O<sub>3</sub>] = 23.2 ± 12.5 ppb {spanning 0-99 ppb}. Through the use of High Dimensional Model Representation (HDMR), we show that interference effects derived from the variable ambient gas-concentration mix and changing environmental conditions over three seasons ([Temperature] = 23.4 ± 8.5°C; {spanning 4.1 to 45.2°C} and [Relative Humidity] = 50.1 ± 15.3% {spanning 9.8-79.9%}) can be effectively modelled for the Alphasense CO-B4, NO-B4, NO<sub>2</sub>-B43F, and Ox-B421 sensors, yielding (5-min average) root mean square errors (RMSE) of 39.2, 4.52, 4.56, and 9.71 ppb respectively. Our results substantiate the potential for distributed air pollution measurements that could be enabled with these sensors.

## 1. Introduction

Protecting populations from exposure to poor air quality is one of the greatest public health challenges, affecting all nations on earth (WHO, 2014). For the past half century, developed countries have made an effort to measure concentrations of major pollutants known to degrade health or damage plants and physical structures. Generally, the focus has been on the most populated areas, with an intent to assess daily, monthly or annual concentrations on a regional basis. While greater spatial and time resolution has been desired, the costs of purchasing and operating

Deleted: Leah R. Williams<sup>1</sup>,

Deleted: , breathe,

Deleted: stakeholders

Deleted: issues

Formatted: Not Superscript/ Subscript

Deleted: .

Formatted: Not Superscript/ Subscript

Formatted: Subscript

Formatted: Subscript

Deleted: and the ambient-gas concentration mix encountered at an urban neighborhood site

Deleted: .

Deleted: improving

Deleted: the credibility of

Deleted: pollutant

Deleted: made

Deleted: the air environment

instruments sufficiently robust, accurate and free of interferences to generate reliable data has been prohibitive – an instrument to assess a single pollutant at ambient levels can cost many tens to hundreds of thousands of US dollars.

In this situation it is therefore easy to understand the motivation to develop inexpensive, rapid-response air quality (AQ) monitoring devices that can be deployed in large numbers around point sources or throughout specific neighborhoods, to create the desired high spatial and temporal resolution AQ data grid (Snyder et al., 2013; Kumar et al., 2015; McKercher et al., 2017). Indeed, within the past decade, researchers, entrepreneurs, and manufacturers have pursued the development, deployment, and evaluation of lower-cost devices that measure air pollution (Mead et al., 2013; Williams, 2014b; Masson et al., 2015; Jiao et al., 2016; Lewis et al., 2016; Castell et al., 2017; Mueller et al., 2017), [Hagan et al., 2017 + Zimmerman et al., 2017](#).

While electrochemical (EC) sensors have formed the basis for workplace and hazardous leak detection applications for many decades (Stetter and Li, 2008), their transition from workplace to ambient air is accompanied by much lower target concentration ranges over which the sensors must accurately measure the analyte species of interest (Borrego et al., 2016). Coincident with the need to resolve much lower concentrations is the need to fully understand and model the influence of non-analyte interferences resulting from changing temperature, humidity, pressure, or other gas molecules that may compete with the oxidation/reduction reactions occurring at the working electrode of a given EC-sensor (Mueller et al., 2017). Unless great care is taken when measuring ambient air pollutants, interferences may result in reported pollutant concentrations that are orders of magnitude greater than the true values. At the core of this quantification challenge is the fact that electrochemical sensors rely on resolving very small changes in current ( $\mu\text{A}$ ), and in turn, reliably converting that raw sensor signal into a concentration. The path from raw sensor output to concentration requires (1) a mechanical design that provides consistent, empirically validated sampling of the ambient air, (2) low-noise electrical circuitry (potentiostats) to amplify and resolve small changes in current, (3) electronic filters to remove electrical transients (e.g., radiofrequency (RF) interference) and (4) a method for converting raw signal to concentration that takes into account calibration and interference data.

In order to calibrate and characterize interferences, laboratory and/or field based co-location experiments must be executed spanning the full range of pollutant concentrations and ambient sampling conditions that may be encountered in an actual stand-alone deployment. Deploying lower-cost AQ sensor systems in the absence of such calibration significantly undermines the credibility of the data. Indeed, reports have appeared recently raising concerns about the reliability of data produced from inexpensive monitoring devices containing EC-sensors (Lewis and Edwards, 2016).

This paper describes results obtained from a newly developed, integrated lower-cost EC-sensor system, ARISense, which has been developed at Aerodyne Research, Inc. for simultaneous, real-time measurement of a wide range of ambient-level atmospheric pollutants and accompanying meteorological metrics. Air quality monitoring systems can be roughly divided into three cost tiers, 1) high cost/high accuracy systems costing tens to hundreds of thousands of US dollars, such as those used at regulatory monitoring stations, 2) lower cost systems costing a few to ten thousand US dollars, such as the ARISense system or recently developed Real-time Affordable Multi-Pollutant (RAMP) package developed by Carnegie Mellon University and SenSevere (Zimmerman et al., 2017), and 3) low cost systems (costing tens to hundreds of US dollars) designed for the consumer market that typically only measure a single pollutant and generally suffer from poor quality data (EPA, 2017). The goal of second tier systems is to provide data quality approaching Tier 1 at a fraction of the cost. In this paper, we describe the mechanical and electronic design of the

Formatted: Highlight

Deleted: the

Deleted: S

Deleted: generate very

Deleted: We will

ARISense system, and demonstrate a field-based calibration technique that combines co-located measurements with a High Dimensional Model Representation (HDMR) of the interferences. Our results show that lower-cost EC-sensor systems can provide reliable measurements of air pollution under real-world ambient concentrations.

## 2. Experimental

### 2.1 ARISense

The ARISense system used in the present study (version 1.0) measures ambient levels of five gaseous pollutants (CO, NO, NO<sub>2</sub>, O<sub>3</sub>, and CO<sub>2</sub>), atmospheric aerosol particles (0.4 – 17 μm in diameter), and related meteorological and environmental parameters (temperature (T), pressure (P), relative humidity (RH), wind speed/direction, solar irradiance, and noise). Mechanical drawings of the instrumented ARISense system are shown in Figure 1. Each ARISense system is housed in a NEMA weather-proof enclosure (Polycase, PN: YH-080804; 21.8 cm L × 13 cm D × 21.8 cm H) weighing approximately 2.7 kg fully integrated. ARISense v1.0 is designed for stationary fixed-site monitoring with access to 120-240V AC power, exterior pole/surface mounting hardware and a consistent sampling orientation relative to the surface.

ARISense v1.0 contained the following EC-sensors (purchased from Alphasense, Ltd.; UK): Carbon monoxide (CO-B4), nitric oxide (NO-B4), nitrogen dioxide (NO<sub>2</sub>-B43F), and total oxidants (Ox-B421). (More recent versions of ARISense have been upgraded to model Ox-B431.) The integrated system also includes a non-dispersive infrared (NDIR) carbon dioxide (CO<sub>2</sub>) sensor (Alphasense Pyro-IRC-A1) and an optical particle counter (OPC) for measurement of particulate matter size distributions (number-count;  $0.4 \leq d_p \leq 17 \mu\text{m}$  over 16 size bins; Alphasense OPC-N2). The following environmental and meteorological measurements are also included: Relative humidity/temperature sensor (Sensirion AG, PN SHT21), barometric pressure/temperature sensor (BOSCH, PN BMP180), solar intensity sensor (OSRAM Opto Semiconductors, PN: BPW 34), and a microphone for audible noise detection (CUI, Inc. PN: CMC-5044PF-A). An anemometer (Davis Instruments, Vantage Pro 6410) for wind speed and direction was mounted to the top of the ARISense NEMA enclosure, measuring conditions ~60 cm above the sampling inlets (see Fig. 1 for reference).

ARISense electronics were designed to integrate all sensor measurements into a unified data acquisition framework and provide user access/control over the system's configuration and operation. EC-sensor signals were collected and processed by custom built electronics designed to minimize noise and amplify raw signals (i.e., potentiostat circuitry). Connectivity for v1.0 systems was enabled via hard-line CAT-5 ethernet connections (Lantronix XPort-Pro). Data was saved at user-defined sampling intervals (5-60s) onto a local USB drive and (if internet-connected) to the ARISense database (<https://arisense.io/>), where data is available for real-time visualization and download. Upgraded ARISense systems configured for cellular connectivity and stand-alone solar power are currently under development.

The ARISense system has two sampling inlets, one for measuring gas-phase pollutants and the other dedicated to the measurement of particulate matter. In both cases, the air flow is driven by small DC-powered fans embedded at the downstream end of the sample flow path, minimizing the loss of sticky or reactive gas molecules (NO<sub>2</sub>, O<sub>3</sub>) or particles due to surface reactivity or deposition. The gas sample flow includes both an intake and an exhaust port in the NEMA enclosure, protected from water penetration via 3D-printed rain hoods (Formlabs; Form 2, Stereolithography 3D printer)

Deleted: called v1  
Deleted: was equipped to  
Deleted: .  
Deleted: .  
Deleted: .  
Deleted: .  
Deleted: ground

mounted to the exterior of the case (see components H, I in Fig. 1). The gas sampling flow manifold and internal PCB mounting brackets were also 3D-printed. Laboratory tests reveal that the 3D-printed material is inert to NO<sub>2</sub> and O<sub>3</sub> and does not result in significant losses of either species when sampling ambient-level concentrations. The gas sampling manifold provides a consistent, compact interface for the 4 electrochemical sensors as well as the CO<sub>2</sub> sensor. The manifold includes an embedded RH/T sensor positioned adjacent to the electrochemical cells which is used to model the temperature and relative humidity-derived interference effects on the raw sensor response. Given the active flow of the gas sampling inlet and minimal residence time (~1s) of the sample air within the manifold, the RH and T measurements recorded by the ARISense system, closely track changes in ambient RH and T conditions. Over the co-location period described here, measurements inside the flow manifold were within 10% of the ambient values even under conditions of direct sunlight. Note that the CO<sub>2</sub> measurements are not discussed in this paper which focuses on the electrochemical sensors, but will be addressed in a future manuscript.

The particle inlet is on the bottom face of the NEMA enclosure (Fig. 1C). Given the body of evidence implicating PM<sub>2.5</sub> concentrations in adverse health outcomes (Lim et al., 2012), recent years have seen substantial growth in the development, evaluation, and deployment of low cost OPCs (Holstius et al., 2014; Williams, 2014a; Han et al., 2017; Zikova et al., 2017). The principal measurement challenge of these devices is the minimum size detection limit, often  $d_p \geq 0.5 \mu\text{m}$  (for devices that cost ~250 to 800 USD) or  $d_p \geq 1.0 \mu\text{m}$  (cost ~15 to 200 USD). Unfortunately, given these size detection limits, such low-cost OPCs are inadequate when the accumulation mode aerosol size distribution peaks at  $d_p \leq 0.25 \mu\text{m}$ , which is typical in most urban locations. Low-cost OPC size detection limits also make near-field particulate combustion emission characterization (i.e., near roadways) very challenging since the combustion mode of particles is typically  $d_p < 0.1 \mu\text{m}$ . A detailed assessment of the ARISense particulate measurements in laboratory and field experiments will be provided in a subsequent manuscript.

This paper presents results for the four electrochemical sensors in a single ARISense system. Note that nominally identical electrochemical sensors can have widely different sensitivities and exhibit variable environmental interference effects, such that the calibration model described in this paper will need to be re-developed for each individual ARISense system.

## 2.2 Measurement site

Two ARISense systems (indicated with yellow circles in Fig. 2) were deployed south of Boston, MA from July to November, 2016. This initial deployment of the ARISense systems was in conjunction with an existing 4-node network (the Dorchester Air Quality Sensor System (DAQSS) project) established in January of 2016. The DAQSS node locations are indicated with green markers on the map. The neighborhoods of Roxbury and Dorchester are among Boston's largest and most economically diverse, including low-income residential areas interspersed with light and heavy industry, as well as the Interstate 93 corridor which runs along the eastern edge of Dorchester. Given their location and activities therein, Dorchester and Roxbury experience a high frequency of automobile, commercial truck, and heavy duty diesel traffic, much of which is constrained to stop-and-go driving patterns on congested, narrow streets, in close proximity to housing and pedestrians. The original DAQSS deployment and initial ARISense proof-of-concept efforts were motivated by the need to assess the viability of lower-cost AQ sensor systems in communities suffering

Deleted:

Deleted: m

Deleted: enclosure

Formatted: Subscript

Deleted: -\$

Deleted: USD

Deleted: \$USD

Deleted: -\$

Deleted: USD

Deleted: \$

Deleted: USD

Deleted: y

Deleted: s

Deleted: sensor and each

Deleted: .

Deleted: . Asthma is a serious concern for residents of Roxbury and North Dorchester. From 2001 to 2010 adult resident asthma rates in North Dorchester doubled from 9 to 18%, whereas city-wide asthma rates in 2010 were 11% (Backus, 2012).

from environmental health knowledge gaps, such as the unexplained doubling of the adult asthma rate in North Dorchester between 2001 and 2010 (Backus, 2012).

In order to validate our approach, each ARISense system was co-located with a Massachusetts Department of Environmental Protection (MA DEP) air quality monitoring station (indicated with red circles on the map) for the duration of the present study. This paper presents ARISense and MA DEP reference data for the Roxbury site (left hand yellow circle in Fig. 2) located adjacent to Harrison Avenue in Dudley Square (latitude: +42.3295 longitude: -71.082619). Forthcoming papers will present results from the DAQSS project and the I-93 ARISense node location, covering lower-cost AQ sensor results over longer deployment timescales (18-24 mo.) and across multiple types of microenvironments in Roxbury and Dorchester.

Deleted: 12-18

### 2.3 Reference data

The MA DEP Roxbury air monitoring site (id: 25-025-0042), established in December, 1998, hosts continuous and semi-continuous gas and particle phase measurements. The reference measurements used in this study include ozone (O<sub>3</sub>, Teledyne Model T400 Photometric Ozone Analyzer), carbon monoxide (CO, Teledyne Model T300/T300M Carbon Monoxide Analyzer), and nitrogen oxides (NO, NO<sub>x</sub>, NO<sub>2</sub>, Teledyne Model T200 Nitrogen Oxide Analyzer). The reference NO/NO<sub>2</sub> measurement is based on chemiluminescence. This method relies on converting NO molecules to NO<sub>2</sub> via exposure to O<sub>3</sub>. Operationally, there are two measurements channels, one for NO alone and one for total NO<sub>x</sub>. In the NO<sub>x</sub> channel, a catalytic-reactive converter is used to convert any existing NO<sub>2</sub> molecules to NO, prior to exposure to O<sub>3</sub>. NO<sub>2</sub> concentrations are determined by taking the difference between NO<sub>x</sub> and NO. Additional on-site reference measurements include a meteorological tower (relative humidity, temperature, pressure, wind direction, wind speed, solar intensity; MetOne), PM<sub>2.5</sub> (BAM, Beta Attenuation Mass Monitor), PM<sub>10</sub>, black carbon, and several off-line gravimetric filter samplers including PM<sub>2.5</sub> speciation. Given its level of instrumentation, the Roxbury location is considered an N-core site within the DEP network of monitoring stations across the state and provides critical data comparisons for determining the viability of lower-cost AQ sensor systems. For the current study, DEP provided real-time (1-minute average) pollutant concentration data files from its reference gas analyzers to permit data comparisons with the ARISense EC-sensor response under rapidly changing conditions of temperature, humidity, and ambient gas concentrations.

Deleted: pressure,

### 2.4 ARISense Calibration

Calibration is a critical issue for trusting the output of EC-sensors. Recent papers (Lewis et al., 2016; Castell et al., 2017) have highlighted that the lack of rigorous calibration protocols for lower-cost AQ sensors results in significant potential error when the sensor system is deployed in ambient conditions. For example, Mead et al. (2013) modelled the temperature dependent baseline drift of an Alphasense NO sensor using an exponential curve fit through 24 hours of ambient data. Their analysis revealed that temperature-derived baseline-drift could exceed a +600 ppb bias, if unaccounted for in their calibration (sampling between 20 and 28 C). Considering that the ambient NO concentration range encountered in the current study was 0 - 200 ppb with temperatures varying from 5 - 45 C (5-min averages), modelling the NO-B4 sensor temperature-derived interference is crucial to obtaining useful measurements from the sensor. As Mead et al. (2013) point out, when measuring gas concentrations at the ppb level, temperature and humidity

interference effects have a first-order impact on quantification, whereas drift in sensitivity over time has second-order effects (much smaller in magnitude than temperature or humidity influence). Both first and second order effects need to be correctly parameterized in order to apply lower-cost sensors to [long-term \(~18-36 mo.\)](#) ambient outdoor air quality measurements.

5 Alphasense provides some guidance to customers regarding calibration and temperature-compensation of electrochemical sensor response (Alphasense Application Note #AAN 803-03, December 2014). This document highlights the utility of including a fourth electrode in their B4-series electrochemical sensors such as were used in this study. The purpose of this fourth electrode (called the auxiliary electrode, AUX) is to provide a real-time correction for environmentally-derived interferences at the working electrode (WE). The AUX electrode is comprised of an identical catalyst to that of the WE and is designed to mimic the WE's response to environmental changes such as temperature, pressure, and humidity. Since the AUX electrode is fully submerged in the electrolyte and directly below the WE, the AUX signal should be blind to the target analyte gas species which readily oxidize or reduce at the WE surface (which is exposed to the air on one side and the electrolyte layer on the other). In an ideal world, a simple subtraction of the current generated at the AUX electrode from the current generated by the WE would provide a signal that is linearly proportional to the target analyte over the full concentration range of interest. Unfortunately, we have found that in practice the AUX electrodes in most sensors are not able to track the changes in the corresponding WE over the nominal operational temperatures of the system. Specifically, at temperatures > 25 C, the AUX electrode response lags that of the working electrode and in some cases (CO-B4, for example), the WE and AUX electrode currents diverge as temperature increases (i.e., the WE current decreases with increasing temperature while the AUX electrode current increases with increasing temperature). In this case, recording just the differential current without correction leads to an increasingly negative concentration error for CO at temperatures above 25 C. Alphasense provides users with a table in which, for each sensor model, the user can identify a correction constant to use to compensate for observed behavior at specified temperature ranges. At temperatures  $\leq 20$  C the Alphasense documentation shows that differential measurements remain fairly stable in comparison to the higher temperature conditions. The Alphasense approach to temperature compensation also requires the use of four static constants for each individual EC-sensor – subtracting specific electronic and zero currents from both the WE and AUX electrodes, prior to calculating the difference. While there are some advantages to the additional information provided by the AUX electrode, at temperatures higher than 25 C, the disparate response between the two electrodes can complicate quantification steps considerably.

In practice, we have found that the manufacturer's recommended WE and AUX electrode corrections do not lead to pollutant concentration values of acceptable accuracy for ambient air analysis. In addition, the EC-sensor response is impacted by other environmental conditions besides temperature, such as relative humidity and the concentrations of other [gas phase](#) species. At the low concentrations present in the atmosphere (10s-100s ppb) characterizing the full [\(multi-dimensional\)](#) interference response is critical to achieving reliable measurements. In this work we demonstrate the use of a multi-dimensional mathematical modelling approach (HDMR) that has the ability to adequately identify and quantify the complex EC-sensor response to multiple environmental variables and interfering gas species simultaneously.

## 2.5 High Dimensional Model Representation (HDMR)

The ARISense system uses [high dimensional model representation \(HDMR\) techniques](#) to convert the raw sensor outputs into units of concentration in parts-per-billion by volume, ppb. HDMR is a numerical method consisting of a general set of quantitative model assessments and analyses for capturing input-output system behavior without reliance on a physics-based model or the sensor manufacturer's empirical correction procedure. When applied to a set of experimental data (with sufficient variability), it can produce a mathematical model relating user-defined input variables to output variables of interest; the resulting model can capture the interdependencies of the variables and provide a mathematical description of the system that is otherwise difficult or impossible to describe with a physics-based model. The HDMR model can be used to identify and quantify which variables and variable interactions have the most impact on the data reduction, [relative to an identified output \(i.e. reference concentration\)](#). In collaboration with the research group of Prof. Herschel Rabitz of Princeton University, Aerodyne has implemented HDMR methods in a software tool called *ExploreHD*, providing graphical and command line user interfaces to HDMR algorithms.

The details of the HDMR algorithms used here are discussed in detail elsewhere (Li and Rabitz, 2010; Sipilä et al., 2010; Li and Rabitz, 2012; Li et al., 2012). One of the key underlying tenets of the HDMR framework is that many input-output relationships for complex physical systems can be captured adequately by [low-order](#) combinations of input variables, even in systems with high-dimensionality in input variables. Each component function provides an additive contribution to the overall model prediction. The modelling process involves three steps. In the first step, the user specifies a maximum variable interaction order (for example, [a 2<sup>nd</sup>-order HDMR model would allow component functions involving combinations of two input variables](#)), and the HDMR algorithm considers orthogonal component functions (in this case, cubic polynomials) involving all possible variable combinations up to the maximum specified order. In the second step, a statistical analysis ([using F-test](#)) is performed to identify the input variables and combinations of input variables that contribute significantly to [variation](#) in the output of interest. In the final step, coefficients for component basis functions are calculated through a least squares analysis that minimizes the deviation between HDMR model prediction and the training data. The coefficients and the associated orthogonal basis functions determined through the above analysis together define an HDMR model for the input-output relationship under consideration.

In the current study, the HDMR approach uses the raw EC-sensor output and environmental variables to model the multi-dimensional [relationship between sensor output and the reference concentration](#). We used approximately 25% of the dataset to train the model. Sensor interference can be a product of the combined influences of temperature, humidity, pressure, non-analyte gas species, etc. The structure of the computational model accounts for both absolute (*i.e.*, highest to lowest concentrations) and transient ( $\Delta x/\Delta t$ ) changes in the sampling conditions encountered by the sensor system. By spanning three seasons in the Northeastern United States, a wide range of environmental conditions was captured within the training window for the model. This emphasizes the advantage (*i.e.*, variability in sampling conditions) and disadvantage (extended time-span) of a field-based co-location approach to sensor calibration. [The HDMR models developed in the current work were 2<sup>nd</sup> order \(examining all possible input-parameter pairs\) with orthogonal polynomial component functions allowed up to degree of three \(cubic\) in each input variable.](#)

[The metrics used to evaluate the model are the slope and intercept of a linear least squares regression of the model output with the reference measurements, the coefficient of determination of the linear fit \( \$r^2\$ \), the root mean square error](#)

Deleted: intricate

Formatted: Font: Italic

Deleted:

Deleted: the

Deleted: the correct

Deleted: n

Deleted: interaction order = 2

Formatted: Superscript

Deleted: any

Deleted: to interact

Deleted: changes

Deleted: .

Deleted: surface

Deleted: 3

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: the third power

Deleted: for each possible input parameter pair (*i.e.* cubic polynomial expressions:  $Y = A + f(x_i+x_j) + f(x_i+x_j)^2 + f(x_i+x_j)^3$ ).

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Superscript

(RMSE), the mean absolute error (MAE), and the mean bias error (MBE). The equations for these metrics are given in Table S1 and model-to-measurement results are summarized in Tables 2 and 3.

An example of how the HDMR model is developed for the NO-B4 sensor is provided in the Supplemental Material. The left column of Table S2 lists all available input parameters and the other columns denote which parameters were included in the input matrix for each model run. The bottom rows list the root-mean-square error (RMSE), mean absolute error (MAE), and mean bias error (MBE) for each model run, displaying statistical performance metrics for both the training data (model generation) and test data (model evaluation).

The data presented in this paper were recorded over a 4.5-month sampling interval (July 7, 2016- November 23, 2016). All four electrochemical sensors used in this study were first removed from their packaging on May 9, 2016. That means that from out-of-package, the sensors had aged ~6.5 months by November 23. The manufacturer quoted lifetime for degradation of the signal to 50% is 36 months for the CO sensor and 24 months for the NO, NO<sub>2</sub> and O<sub>2</sub> sensors. Given that these lifetimes are significantly longer than the deployment time scale analysed here, we did not include a time-dependent sensitivity term in the input matrix of our HDMR model runs. The results presented here therefore assume that the sensitivity of each of the electrochemical sensors did not appreciably drift over the 4.5 month deployment. In subsequent studies we will analyze sensor response over longer deployment timescales (18 to 24 mo.) to investigate the importance of including a time-based parameter to track and correct for drift in sensor response with time.

### 3. Results and Discussion

#### 3.1 ARISense meteorological/environmental data

Continuous 5-min average non-pollutant data acquired with the ARISense system is shown in Fig S2 of the supplemental, tracking ambient variability in temperature, pressure, humidity, solar intensity, ambient noise, wind speed, and wind direction at the Roxbury DEP monitoring site. The total sampling timespan covers the transition from mid-summer through late fall in the Northeastern United States (July through November), with meteorological conditions changing from warmer and more humid to cooler and less humid. The ARISense system ran continuously throughout the sampling interval with the exception of a ~ 1-week period during which the node was physically removed from the site for a separate experiment. The directionality of the wind fields at this site is predominantly from the N to NW (red-maroon) with occasional NE flow (blue-purple). Temperature and humidity measurements shown reflect the conditions within the gas-sampling flow-cell of the integrated system, characterizing the environmental conditions at the surface of the electrochemical sensors. Such environmental measurements are critically important for reconciling the interference effects of ambient conditions, especially humidity (water concentration) and temperature, on the raw signal from each electrochemical cell.

#### 3.2 ARISense electrochemical sensor data

Figure 3 shows the time-series for a ~ 72-hour period for the relative humidity, dew point temperature (panel a, solid and dashed lines, respectively), temperature (grey shaded area), and raw differential sensor output (dashed line), reference measurement (thick red line) and model output (thin solid line) for the four electrochemical sensors. The raw

Formatted: Not Highlight

Deleted: EC-

Deleted: gives

Deleted: the possible

Deleted: show

Deleted: a series of

Deleted: s

Deleted: give

Deleted: model run for

Deleted: for each model applied to the

Deleted: (i.e., the remainder of the ambient data excluding the data used for training)

Deleted: Step 1 was to specify a maximum interaction order of 2, meaning that only pairwise interferences are considered. Step 2 was to run the model with all possible inputs (Run 12 in Table S2) and examine the contribution of each pair to minimizing the residual between the model output and the reference measurements. Figure S1 shows the fractional contribution for the pairs that were included in the final model (Run 8 in Table S2); pairs that had a fractional contribution less than 0.1% were excluded. The metrics in Table S2 show that while the model with all possible parameters had the best performance on the training data, it had worse performance on the test data than the final model. This demonstrates the importance of Step 2 in determining the most relevant parameter pairs and in excluding pairs that have negligible impact on the fit to the training data. ¶

Formatted: Not Highlight

Deleted: It is important to recognize that the data presented in this paper were recorded over a 4.5-month sampling interval (July 7, 2016- November 23, 2016). All four electrochemical sensors used in this study were first removed from their packaging on May 9, 2016. That means that from out-of-package, the sensors had aged ~6.5 months by November 23. The manufacturer quoted lifetime for degradation of the signal to 50% is 36 months for the CO sensor and 24 months for the NO, NO<sub>2</sub> and O<sub>2</sub> sensors. For the purposes of this study since these lifetimes are at least 5 times longer than the deployment, we are assuming that the sensitivity of each of the electrochemical sensors did not appreciably drift over this time interval the 4.5 month deployment. This means that the HDMR model built in this case does not take sensor age into account. In subsequent studies we will analyze sensor response over longer deployment timescales (18 to 24 months) to investigate the importance of including a time-based parameter to track and correct for drift in sensor response with time.¶

Formatted: Subscript

Formatted: Subscript

Formatted: Highlight

Deleted: 1

Deleted: precise

Deleted: day

Deleted: Time-series of 5-min average raw differential signal ( $\Delta mV$  = working electrode - auxiliary electrode) obtained from each electrochemical sensor (CO, NO, NO<sub>2</sub>, and O<sub>2</sub>) in the ARISense system are shown in Fig. 3.

Deleted:

differential sensor output (diff mV) is displayed as a voltage which is linearly proportional to the difference in current generated within the electrochemical cell at each electrode (working and auxiliary). The correlation plots between the raw EC-sensor output and the reference measurements are shown in Figs. 4a to d, with each data point colored by flow-cell temperature. The intercept, slope and  $r^2$  for the linear regression indicated with a red line are given in Table 1.

The raw differential signals obtained from the CO-B4 sensor track reasonably well with the CO concentrations measured by the co-located DEP monitor (Fig. 3b and 4a), demonstrating the relatively small influence of ambient temperature, humidity or other chemical species on this EC-sensor. The NO sensor raw output also tracks reasonably well with the reference measurements (Fig. 3c and 4b) except at temperatures over 25 C when the EC-sensor overestimates NO by a factor of 2 to 3 compared to lower temperatures. This suggests that the temperature dependence of the working and auxiliary electrodes in this NO sensor do not track one another at sample temperatures > 25 C, and that additional temperature-correction is necessary to obtain reasonable NO concentrations from the raw sensor outputs.

The NO<sub>2</sub> and O<sub>3</sub> raw sensor outputs track less well with the reference measurements ( $r^2 < 0.2$  in Figs. 4c and d). The differential NO<sub>2</sub>-B43F sensor response (Fig. 3d and 4c) indicates a strong temperature dependence that is not compensated for by the auxiliary electrode, suggesting that additional temperature compensation algorithms could improve the result. The differential signal from the Ox-B421 electrode shows poor correlation with the reference data overall (Fig. 4d). There is some temperature-dependence, but the additional variation suggests that other factors play an important role. The Ox-B421 sensor is comprised of the same catalyst (working and auxiliary electrode material) as the NO<sub>2</sub>-B43F, and is therefore-sensitive to NO<sub>2</sub> in addition to O<sub>3</sub>. The key difference between these two sensors is the presence of an O<sub>3</sub>-scrubbing filter upstream of the working electrode in the NO<sub>2</sub>-B43F sensor package. Laboratory results indicate that the Ox-B421 sensor is ~2x more sensitive to NO<sub>2</sub> than to O<sub>3</sub> molecules.

As Fig. 3 shows, the magnitude of the interference signal due to temperature alone (for NO, NO<sub>2</sub>, and O<sub>3</sub>) can easily mask real variation in pollutant concentrations. The raw signal behavior observed for each sensor type is different, underscoring the necessity of species-specific HDMR models to reconcile each sensor type's characteristic interferences. In addition, substantial (~ 2-3x) differences (in sensitivity and baseline) exist for batches of nominally identical sensors measuring the same concentration. Therefore, the HDMR models built for a given integrated system are specific to a given set of sensors, and must be generated for each system separately to achieve reliable concentration data. Within the framework of an individual ARISense system, 4 distinct HDMR models are built, one for each EC-derived pollutant species of interest.

### 3.3 HDMR Analysis

The training data for the HDMR model were chosen to provide comprehensive coverage of environmental variability spanning the July-November sampling interval. It was important to include (1) sensor responses to the range of gas concentrations encountered in ambient air (near-zero to high concentration transient spikes in pollution), (2) the range of temperatures and various rates-of-change in temperature, and (3) the range of measured water content of the sample air in the flow-cell. The goal was to include a wide enough range of training data to avoid extrapolation errors when applying the model to the test dataset (all ambient co-location data not included in the training dataset). Figure S3 shows the distributions of temperature, reference measurement, dew point temperature and relative humidity for the training data for the CO-B4 HDMR model, overlaid with corresponding distributions of the test data. We did not

- Deleted: signal
- Deleted: related
- Deleted:
- Deleted: The time-series of co-located reference concentrations are plotted on the right axes for each gas-phase pollutant of interest.
- Deleted: s
- Deleted: panels a-d of
- Formatted: Superscript
- Deleted: correlate
- Deleted: 4a
- Deleted: correlates
- Deleted: ,
- Deleted: 4b
- Deleted: alone
- Deleted: correlate
- Formatted: Font: Italic
- Formatted: Superscript
- Deleted: 4c
- Deleted: residual
- Deleted: ¶  
In order to demonstrate the varying effect of temperature on the different electrodes, Figure 5 shows raw 2-min average output currents for both the working (darker hue) and auxiliary (lighter hue) electrodes for each electrochemical sensor over a 48-hour time interval. Reference concentrations of each pollutant species are plotted (as red dashed-lines) on the right-hand axes for comparison, and the flow-cell temperature is displayed as a filled histogram in the background of the time-series. Temperature changes in excess of 22 C are observed in as little as 10 hours. Offsets are observed between the working and auxiliary electrodes of all four sensors, but the extent to which the working/auxiliary pair track one another as environmental conditions change is highly sensor-dependent. The CO-B4 sensor appears to be fairly insensitive to the temperature changes encountered, with a positive-going differential reflecting changes in the ambient concentration of CO. In contrast to the muted temperature response of the CO sensor, the NO-B4 sensor shows a strong temperature dependence in both the working <... [1]
- Deleted: ure
- Deleted: 5
- Deleted: derived from
- Deleted: ,
- Deleted: t-derived variation
- Deleted: unique
- Deleted: ¶
- Deleted: The next step in the analysis was to determine wh<... [2]
- Deleted: time intervals
- Deleted: over which the model input matrix was generated
- Deleted: rest of the
- Deleted: and
- Deleted: for the
- Deleted:
- Deleted: (remainder of the 4.5 month deployment, excludi<... [3]

attempt to minimize the amount of ambient data used for training, or vary the timing of the training data with respect to the test data. Approximately 25% of the full time series was used to generate the model (Table 2 and indicated with grey bars in Fig. 5). The exact fraction of data used for training was slightly different for each sensor due to differing calibration schedules for the reference measurements (which automatically excludes sensor data from the training or test datasets). For each sensor, the set of inputs included in the input data matrix was optimized as described in Section 2.5 and the Supplemental Materials.

Correlation plots of model-derived pollutant concentrations and reference concentrations for the training data are shown in the middle panels (e-h) of Fig. 4. The linear regression fit (solid red line) and a 1:1 line (dashed black line) are shown for all species, and all data points are colored by flow-cell temperature. The performance metrics are presented in Table 2. The lack of a temperature-dependent rainbow in the scatter plots shown in Fig. 4e-h (with the exception of O<sub>3</sub>, for which ambient concentrations are expected to be temperature-dependent) indicates that the model has effectively compensated for the variable temperature dependent response of the working and auxiliary electrodes within each cell. The remaining scatter in the correlation plots is random noise attributed to the electrodes themselves and the electronics. The high correlation coefficients ( $r^2=0.94-0.96$ ) for CO, NO and NO<sub>2</sub> indicate that, when trained appropriately, the HDMR model provides improved compensation for the environmental interferences that complicate interpretation of raw EC-sensor outputs. The much lower correlation coefficient ( $r^2=0.65$ ) for O<sub>3</sub> suggests that additional parameters may be needed to fully explain the behaviour of this EC sensor.

The HDMR models were then used to analyse the remaining ~ 75% of the data (the test set). The correlation plots for the model output versus the reference measurement for the test data are shown in Figs. 4i-l and the performance metrics are presented in Table 3. Figure 5 shows the time-series for the 5-min averages of the modelled (sensor) and reference gas concentrations, with the training data intervals indicated with grey bars. The high correlation coefficient for CO and NO ( $r^2 > 0.8$ ) and moderate correlation coefficient for NO<sub>2</sub> ( $r^2 = 0.69$ ) indicates the strength of the model at capturing the ambient variability in pollutant concentrations encountered at the site, despite wide variations in ambient temperature and humidity over the changing seasons. The higher scatter in the O<sub>3</sub> correlation plot, and correspondingly low  $r^2 = 0.39$ , might be due to the fact that O<sub>3</sub> is obtained by training the Ox-B421 sensor output to reference O<sub>3</sub>; the 2:1 sensitivity ratio for NO<sub>2</sub> vs. O<sub>3</sub> of the Ox-B421 means that the variability in ambient NO<sub>2</sub> concentrations adds considerable noise to the Ox-B421 sensor signal. The input matrix for the Ox-B421 HDMR model includes the raw data captured with the NO<sub>2</sub>-B43F sensor, but the inclusion of this additional information only marginally improves the reduction of the Ox-B421 data to O<sub>3</sub> concentration. It should be noted that the Ox-B421 sensor is not the latest version released by Alphasense and improvements may be realized with the design of their most recent model (Ox-B431). An important design considerations pertaining to the Ox sensor is that its response time closely mimics that of the NO<sub>2</sub> sensor. With this time-response aligned, the cross-sensitivity of the Ox sensor to NO<sub>2</sub> can be effectively subtracted from the raw (Ox) sensor signal. If the time response of the two sensors is different, reconciling O<sub>3</sub> concentrations from the Ox sensor is far more challenging. These considerations highlight the iterative and rapidly evolving nature of low-cost AQ sensor components. Lower-cost air quality sensor quantification will likely improve over the coming years through advances at both the component manufacturer level (e.g., Alphasense Ltd, improving materials chemistry/catalyst and sensor-design) and system integrator level (e.g., Aerodyne Research, Inc, further developing ARISense HDMR interference modelling through laboratory and field-based measurements).

- Deleted: In total 35
- Deleted: data
- Deleted: ere
- Deleted: in
- Deleted: and are
- Deleted: 6
- Deleted: by systematically testing all possible input param (... [4])
- Deleted: The
- Deleted: s
- Deleted: (black dashed lines)
- Deleted: quite
- Deleted: we have been able to
- Deleted: different electrochemical sensors, and the workin (... [5])
- Deleted: plus some inherent scatter in the reference data
- Deleted: R<sup>2</sup>
- Deleted: 76
- Formatted (... [6])
- Deleted: data reduction
- Deleted: dramatically
- Formatted (... [7])
- Commented [LW3]: Of course this begs the question of what.
- Deleted: 6
- Deleted: plots for the full dataset analyzed with the HDMR (... [8])
- Deleted: Correlation plots of sensor-derived and reference (... [9])
- Deleted: robustness of the linear correlation plots
- Formatted (... [10])
- Formatted (... [11])
- Formatted (... [12])
- Formatted (... [13])
- Formatted (... [14])
- Deleted: can be largely
- Deleted: attributed
- Deleted: with
- Deleted: ,
- Deleted: possible
- Deleted: in
- Deleted:
- Deleted: i
- Deleted: s fact
- Deleted: s
- Formatted (... [15])
- Deleted: ,
- Deleted: calibration
- Formatted (... [16])
- Deleted: ,

5 Closer examination of the model output for 72-hours of the test data in Fig. 3 gives additional clues for improving the model. In Fig. 3d at ~18:00 on 11/2/2016, the model NO<sub>2</sub> exceeds the reference NO<sub>2</sub> by a factor of ~2 during a period of rapidly decreasing temperature and increasing RH. This underscores the importance of the rate of change of input parameters may be important in the model, in addition to the absolute values. Fig. 3e also suggests that the HDMR model for O<sub>3</sub> struggles during times of rapidly changing temperature, particularly when the O<sub>3</sub> concentration is low (< 3 ppb). Future development of HDMR models to support ARISense quantification will include derivatives of key variables as inputs.

10 While Figures 4 and 5 illustrate that the system is capable of determining valid gas phase concentrations across a wide range of environmental variability in temperature, RH, and absolute concentrations, it does not speak to the longer-term stability of the sensors (e.g., how much does the baseline and sensitivity of each electrochemical sensor change with time). However, it should be noted that sensor aging cannot have had a major impact on the data reported here or it would have been impossible for the HDMR model to converge this well without including electrode age as one of the input variables. For the models developed in this work, each data point for each variable had equal weight, whether it was at the beginning, middle or end of the 4.5 month deployment. It is to be expected that aging of EC-

15 sensors will change their sensitivities, due to electrolyte evaporation or dilution, entrapment of contaminants, and repeated exposure to wide swings in T or RH. It will be important to establish the time span over which a given set of EC-sensors (and the HDMR model of that sensor set) can be expected to return reliable pollutant concentration values, using a longer duration (18-24 Mo.) ambient data set; such a study is in progress.

#### 4. Conclusion

20 This study demonstrates that lower-cost air quality sensor systems can adequately characterize ambient urban pollution concentrations on rapid (5-min) timescales, underscoring the potential of integrated sensor systems to add a highly resolved local AQ data-layer to existing pollutant monitoring infrastructure. The ARISense system is a first step toward understanding the extent to which quantification efforts can yield useful results from such systems. Training electrochemical sensor measurements of CO (231±116 ppb; {spanning 84-1706 ppb}), NO (6.1±11.5 ppb; {spanning 0-209 ppb}), NO<sub>2</sub> (11.7±8.3 ppb; {spanning 0-71 ppb}), and O<sub>3</sub> (23.2±12.5 ppb {spanning 0-99 ppb}) with a High Dimensional Model Representation (HDMR) method provided 5-min average RMSE values of 39.2, 4.52, 4.56, and 9.71 ppb for CO, NO, NO<sub>2</sub>, and O<sub>3</sub> respectively. Results indicate that HDMR can effectively model interference effects derived from the variable ambient gas-concentrations in an urban setting and changing environmental conditions encountered over three seasons in the Northeastern United States ([Temperature] = 23.4 ± 8.5C; {spanning 4.1 to

25 45.2C) and [Relative Humidity] = 50.1 ± 15.3% {spanning 9.8-79.9%}).

30 Referring back to the map displayed in Figure 2, it is striking to consider that only 4 official monitoring stations exist within the Boston-metro area (pop. ~700,000). With regard to the Roxbury DEP site 1-minute average reference data, it is important to note that 1-min data files are not typically reported or accessible from regional air quality monitoring sites. Instead, pollutant concentrations are usually reported on 1, 8, or 24-hour averages in accordance with the operational constraints of the measurement device and relevant air quality regulations being enforced. This sampling paradigm is consistent with the regional focus of federal and state monitoring goals, and financial constraints

- Deleted: E
- Deleted: 3 days
- Deleted: around 6
- Deleted: pm
- Formatted: Subscript
- Formatted: Subscript
- Deleted: s
- Deleted: suggests
- Deleted: that
- Formatted: Subscript
- Formatted: Subscript
- Deleted: very
- Deleted: exploration
- Deleted: the
- Deleted: development
- Deleted: of
- Deleted: the parameters as
- Deleted: 6
- Formatted: Font: Italic
- Deleted: entering
- Deleted: In the actual analysis
- Deleted: time
- Deleted: span
- Deleted: training
- Deleted: empirical
- Deleted: and have

Deleted:

Deleted:

imposed due to the expense of the instrumentation and operating costs of a given AQ monitoring station. But as one considers the pollutant sources that [contribute to their local area](#), disproportionate pollution impacts emerge in some neighborhoods more than others. Across urban landscapes, air pollution is inherently heterogeneous, subject to sharp concentration gradients over fast (sub-minute) and short (100's of meters) scales. In order to establish a more rigorous assessment of such disparate impacts, distributed sensor networks are needed to achieve high enough spatial resolution to inform intra-neighborhood differences in air quality. Through such advances, researchers, regulators, and community members can improve their understanding of the pollutant sources that [dictate their local AQ](#). As sensor technologies (and calibration/modelling efforts) continue to improve, the local AQ data layer could play a key a role toward empowering environmental justice advocates to initiate change [and improve environmental public health](#).

**Deleted:** can impact

It cannot be overstated that EC-sensor systems such as ARISense can return reliable data only if calibrated over the full range of pollutant concentrations and meteorological parameters that will be encountered when they are deployed. In the present study, co-location of the ARISense system with the MA DEP reference monitors, coupled with variability of natural [processes](#) and anthropogenic activities, supplied the necessary range of [conditions](#) over the [4.5-month](#) span of the study. In the future, we expect to compress that training period, using a controlled-environment laboratory chamber and mixes of calibration gases representative of the pollutants encountered under ambient conditions. This compression of the training period is especially important when addressing the challenges of sensor-to-sensor variability, finite (< [24-36 mo.](#)) sensor lifetime, and premature damage or failures that will require rapid replacement/re-training of integrated systems.

**Deleted:** have a disproportionate influence on

**Deleted:** calibration events

**Deleted:** four

**Deleted:** into one week

### Acknowledgements

The authors would like to thank the staff at the Massachusetts Department of Environmental Protection for their support, including access to the DEP monitoring stations and raw 1-min data from the reference gas analyzers used in the current study. We are especially grateful to John Lane, Leslie Collyer, Emmy Andersen, Patrick Shea, and Thomas McGrath at MA DEP. ARISense builds upon two AQ sensor development efforts that originated at MIT: the CLAIRITY project and DAQSS network. ESC acknowledges the ground breaking contributions of the Senior Class of '14 Civil and Environmental Engineering Department: Paula Gonzalez, Sidhant Pai, David Ogutu, Linda Seymour, Katherine Spies, Zachary Balgobin, Timothy Wilson, Maria Cassidy, Carolina Kaelin, Priscilla Soto, Hao-Yu Derek Chang, Theresa Santiano-McHatton, Matthew Monheit, Catherine Cheng, Sydney Beasley, Obinna A Okwodu, Vanya Britto, Carmen Castanos, Sharon M Small, and Daphne Basangwa. ESC thanks Colette Heald and Jesse Kroll for their help re-developing the capstone course that resulted in the CLAIRITY network and David Hagan for his support of the DAQSS project (electronics, dev, and system integration) and for helpful discussions regarding advanced modelling approaches to sensor quantification. ESC thanks Ann Backus and Gary Adamkiewicz for their collaboration on the DAQSS project (NIEHS grant P30 ES000002 via the Community Outreach and Engagement Core (COEC) of the HSPH NIEHS-Center for Environmental Health). ESC also thanks Stephen Prescott for his assistance with the ARISense mechanical-electrical assembly, Xavier Cabral for his contributions to the initial electrical design, and Conor Mackinson and Wade Robinson for their mechanical design contributions to the ARISense package.

#### **Deleted: Acknowledgements¶**

The authors would like to thank the staff at the Massachusetts Department of Environmental Protection for their support, including access to the DEP monitoring stations and raw 1-min data from the reference gas analyzers used in the current study. We are especially grateful to John Lane, Leslie Collyer, Emmy Andersen, and Thomas McGrath at MA DEP. ESC also thanks Stephen Prescott for his help with the ARISense mechanical-electrical assembly, Xavier Cabral for his contributions to the electrical design, and Conor Mackinson and Wade Robinson for their mechanical design contributions to the ARISense project.

**Deleted:** ARISense builds on the groundwork laid during the CLAIRITY project at MIT and ESC acknowledges contributions by Jesse Kroll, David Hagan, and the students in the Department of Environmental Engineering Senior Capstone Project in 2014.

## References

- Backus, A. T., K.; Wool, J.; Straubel, M.: Health, Environment, Community - A Tour of Dorchester and Surrounding Neighborhoods for Asthma Prevention and Health Equity, Harvard School of Public Health, NIEHS Center for Environmental Health, and GreenDorchester, 2012.
- 5 Borrego, C., et al.: Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, *Atmospheric Environment*, 147, 246-263, 2016.
- Castell, N., et al.: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, *Environment International*, 99, 293-302, 10.1016/j.envint.2016.12.007, 2017.
- EPA: Evaluation of Emerging Air Pollution Sensor Performance, <https://www.epa.gov/air-sensor-toolbox/evaluation-emerging-air-pollution-sensor-performance>, 2017.
- 10 Han, I., et al.: Feasibility of using low-cost portable particle monitors for measurement of fine and coarse particulate matter in urban ambient air, *Journal of the Air & Waste Management Association*, 67, 330-340, 2017.
- Holstius, D. M., et al.: Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California, *Atmospheric Measurement Techniques*, 7, 1121-1131, 2014.
- 15 Jiao, W., et al.: Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, *Atmospheric Measurement Techniques*, 9, 5281, 2016.
- Kumar, P., et al.: The rise of low-cost sensing for managing air pollution in cities, *Environment International*, 75, 199-205, 2015.
- Lewis, A. C., and Edwards, P. M.: Validate personal air-pollution sensors, *Nature*, 535, 29-31, 2016.
- 20 Lewis, A. C., et al.: Evaluating the performance of low cost chemical sensors for air pollution research, *Faraday Discussions*, 189, 85-103, 2016.
- Li, G., and Rabitz, H.: D-MORPH regression: application to modeling with unknown parameters more than observation data, *Journal of Mathematical Chemistry*, 48, 1010-1035, 2010.
- Li, G., and Rabitz, H.: General formulation of HDMR component functions with independent and correlated variables, *Journal of Mathematical Chemistry*, 50, 99-130, 2012.
- 25 Li, G., et al.: D-MORPH regression for modeling with fewer unknown parameters than observation data, *Journal of Mathematical Chemistry*, 50, 1747-1764, 10.1007/s10910-012-0004-z, 2012.
- Lim, S., et al.: A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010, *Lancet*, 380, 2224-2260, 2012.
- 30 Masson, N., et al.: Quantification method for electrolytic sensors in long-term monitoring of ambient air quality, *Sensors*, 15, 27283-27302, doi:27210.23390/s151027283, 2015.
- McKercher, G. R., et al.: Characteristics and applications of small, portable gaseous air pollution monitors, *Environmental Pollution*, 223, 102-110, 10.1016/j.envpol.2016.12.045, 2017.
- 35 Mead, M., et al.: The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks, *Atmospheric Environment*, 70, 186-203, 2013.
- Mueller, M., et al.: Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich, *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2017-5122, 2017, 2017.
- Sipilä, M., et al.: The role of sulfuric acid in atmospheric nucleation, *Science*, 327, 1243-1246, 2010.
- 40 Snyder, E. G., et al.: The Changing Paradigm of Air Pollution Monitoring, *Environmental Science & Technology*, 47, 11369-11377, 2013.
- Stetter, J. R., and Li, J.: Amperometric Gas Sensors A Review, *Chemical Reviews*, 108, 352-366, 10.1021/cr0681039, 2008.
- WHO: Burden of disease from Ambient Air Pollution for 2012, World Health Organization, Geneva, [http://www.who.int/entity/phe/health\\_topics/outdoorair/databases/AAP\\_BoD\\_results\\_March2014.pdf](http://www.who.int/entity/phe/health_topics/outdoorair/databases/AAP_BoD_results_March2014.pdf), 2014.
- 45 Williams, R. K., A; Hanley, T.; Rice, J.; Garvey, S.: Evaluation of Field-deployed Low Cost PM Sensors, U.S. Environmental Protection Agency, Washington, DC, 2014a.
- Williams, R. L., R., Beaver, M.; Kaufman, A.; Zeiger, F.; Heimbinder, M.; Hang, I.; Yap, R.; Acharya, B.; Ginwald, B.; Kupcho, K.; Robinson, S.; Zaouak, O.; Aubert, B.; Hannigan, M.; Piedrahita, R.; Masson, N.; Moran, B.; Rook, M.; Heppner, P.; Cogar, C.; Nikzad, N.; Griswold, W.; : Sensor Evaluation Report U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-14/143, 2014b.
- 50 Zikova, N., et al.: Evaluation of new low-cost particle monitors for PM 2.5 concentrations measurements, *Journal of Aerosol Science*, 105, 24-34, 2017.
- Zimmerman, N., et al.: Closing the gap on lower cost air quality monitoring: machine learning calibration models to improve low-cost sensor performance, *Atmos. Meas. Tech. Discuss.*, 2017, 1-36, 10.5194/amt-2017-260, 2017.
- 55

Deleted: ¶

Page Break

## Tables

**Table 1. Performance metrics for raw sensor output versus reference measurements.–**

Sensor	N <sub>data pnts</sub>	Y-int (mV/ppb)	Slope	r <sup>2</sup>
CO-B4	29507	3.26	0.25	0.78
NO-B4	33310	6.32	0.30	0.21
NO <sub>2</sub> -B43F	33363	-27.4	0.29	0.18
<b>O<sub>x</sub>-B421</b>	<b>34077</b>	<b>-131</b>	<b>-0.48</b>	<b>0.12</b>

5

**Table 2. Performance metrics for model output versus reference measurements for training data.**

Sensor	N <sub>data pnts</sub>	Y-int (ppb)	Slope	r <sup>2</sup>	RMSE (ppb)	MAE (ppb)	MBE (ppb)	<i>f<sub>train</sub></i> <sup>*</sup> (%)
CO-B4	7974	9.19	0.96	0.96	25.4	16.7	0.02	27.0%
NO-B4	7974	0.46	0.94	0.94	3.37	2.17	-0.01	23.9%
NO <sub>2</sub> -B43F	7874	0.80	0.94	0.94	2.29	1.73	-0.05	23.6%
<b>O<sub>x</sub>-B421</b>	<b>9071</b>	<b>8.34</b>	<b>0.65</b>	<b>0.65</b>	<b>8.24</b>	<b>6.22</b>	<b>0.03</b>	<b>26.6%</b>

\* $f_{train} = N_{training}/N_{total} \times 100$ **Table 3. Performance metrics for model output versus reference measurements for test data (5-min average temporal resolution).**

Sensor	N <sub>data pnts</sub>	Y-int (ppb)	Slope	r <sup>2</sup>	RMSE (ppb)	MAE (ppb)	MBE (ppb)
CO-B4	21533	3.98	0.94	0.88	39.2	24.8	-10.4
NO-B4	25356	1.29	0.94	0.84	4.52	2.83	0.97
NO <sub>2</sub> -B43F	25489	3.26	0.81	0.69	4.56	3.45	1.20
O <sub>x</sub> -B421	25006	13.1	0.47	0.39	9.71	7.34	0.78

15

20

25

5 **Table 4. Comparisons to published results utilizing integrated multi-pollutant systems comprised of**

**Alphasense electrochemical sensors**

Studies	Temporal Resolution (min)	N <sub>data_pts</sub>	Slope	r <sup>2</sup>	RMSE (ppb)	MAE (ppb)	MBE (ppb)
<b>CO-B4 SENSOR</b>							
Jiao et al., 2016 <sup>1</sup>	60	2640-2664 <sup>#</sup>	7.99E-4-8.09E-4	0.63-0.68	NR	NR	NR
Castell et al., 2017 <sup>2</sup>	15	6912 <sup>#</sup>	NR	0.36	170.99	NR	-147.21
Zimmerman et al., 2017 <sup>3</sup>	15	3936 <sup>#</sup>	0.86	0.91	NR	38	0.1
<b>This Work<sup>4</sup> (CO-B4)</b>	<b>5</b>	<b>21533</b>	<b>0.94</b>	<b>0.88</b>	<b>32.9</b>	<b>24.8</b>	<b>-10.4</b>
<b>NO SENSOR</b>							
Jiao et al., 2016	60	2640-2664 <sup>#</sup>	0.883-0.892	0.77-0.87	NR	NR	NR
Castell et al., 2017	15	6912 <sup>#</sup>	NR	0.74	16.35	NR	-0.54
<b>This Work (NO-B4)</b>	<b>5</b>	<b>25356</b>	<b>0.94</b>	<b>0.84</b>	<b>4.52</b>	<b>2.83</b>	<b>0.97</b>
<b>NO2 SENSOR</b>							
Jiao et al., 2016	60	2640-2664 <sup>#</sup>	NR	0.02-0.10	NR	NR	NR
Castell et al., 2017	15	6912 <sup>#</sup>	NR	0.24	30.27	NR	13.30
Zimmerman et al., 2017	15	2304 <sup>#</sup>	0.64	0.67	NR	3.48	-0.4
<b>This Work (NO2-B43F)</b>	<b>5</b>	<b>25489</b>	<b>0.81</b>	<b>0.69</b>	<b>4.56</b>	<b>3.45</b>	<b>1.20</b>
<b>Ox SENSOR</b>							
Jiao et al., 2016	60	2640-2664 <sup>#</sup>	NR	0.15-0.20	NR	NR	NR
Castell et al., 2017	15	6912 <sup>#</sup>	NR	0.29	22.20	NR	6.76
Zimmerman et al., 2017	15	3648 <sup>#</sup>	0.82	0.86	NR	3.36	-0.14
<b>This Work (Ox-B421)</b>	<b>5</b>	<b>25006</b>	<b>0.47</b>	<b>0.39</b>	<b>9.71</b>	<b>7.34</b>	<b>0.78</b>

NR = not reported in manuscript

<sup>#</sup>N calculated assuming 100% duty cycle over specified days of co-location for each study

<sup>1</sup>Results obtained from 2 AQMesh integrated sensor systems (Gen. 3) deployed in Decatur, Georgia US

10 <sup>2</sup>Statistical metrics correspond to average of 24 co-located AQMesh systems deployed in Kirkeveien, Norway

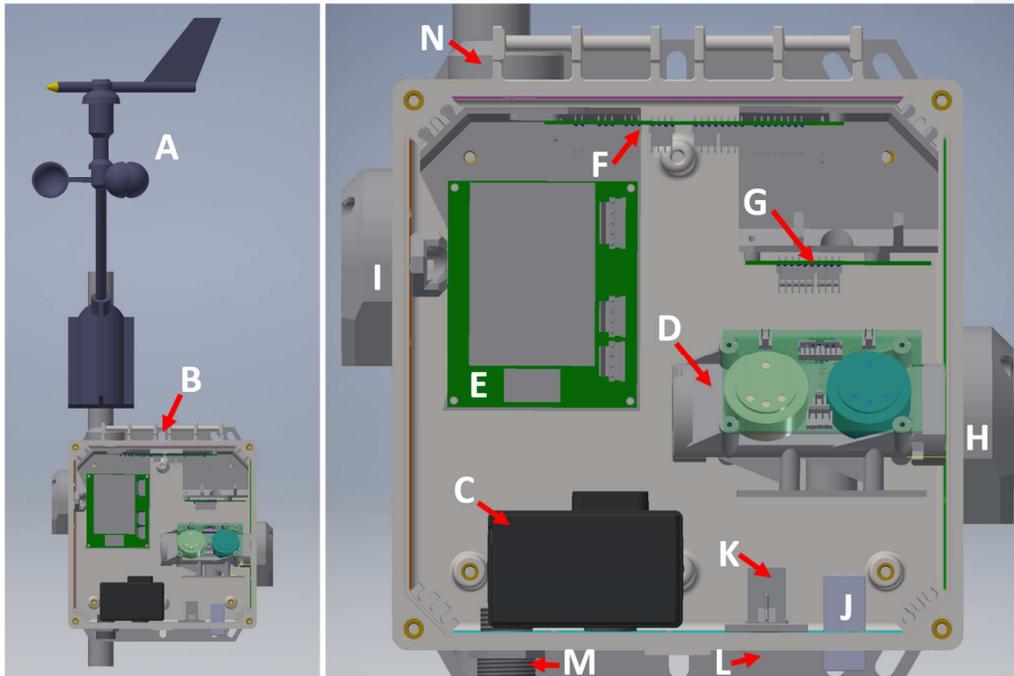
<sup>3</sup>Average co-location test results from 19 Real-time Affordable Multi-Pollutant (RAMP) systems co-located in Pittsburgh, Pennsylvania US

15 <sup>4</sup>Single ARISense system deployed in Dorchester, Massachusetts US

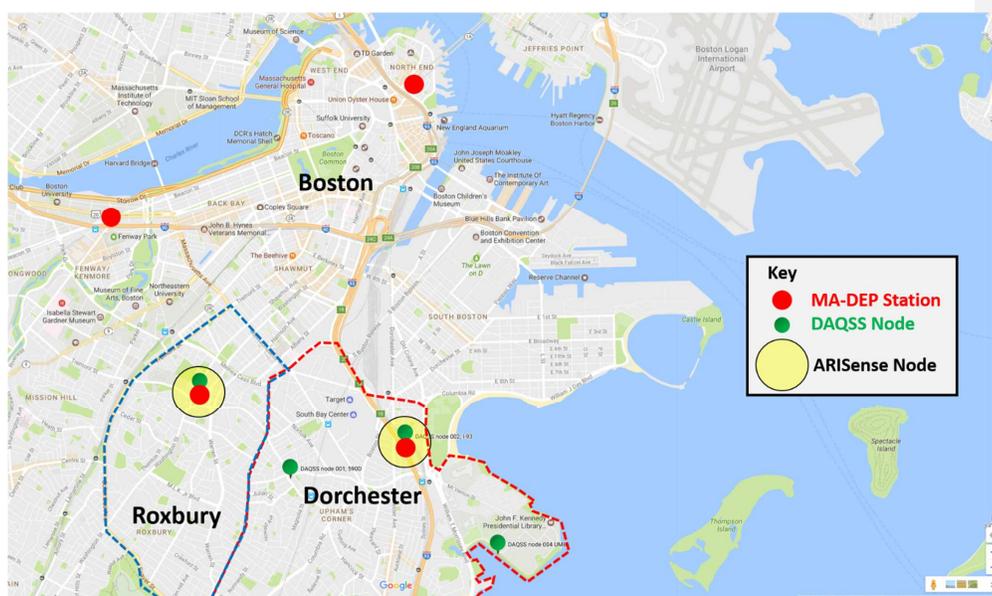
**Deleted:** Table 4. Comparisons to published results utilizing Alphasense electrochemical sensors ¶

... [17]

Figures



5 **Figure 1.** Mechanical drawings (wires excluded) showing the main components of the ARISense system. Each system includes an anemometer (A) mounted to the back-bracket of the NEMA enclosure providing a description of the wind-fields in the immediate proximity to the gas and particle sampling inlets of the system. Mounting brackets for wall or pole-mount configurations attach at position B. Expanded view of the internal components reveals the Optical Particle  
10 Counter (C), gas sampling manifold (D) with embedded electrochemical and NDIR and RH/T sensors, transformer/power PCB (E), main controller PCB (side view) (F), communication PCB for ethernet connectivity (G), gas sampling inlet and exhaust rain hoods (H, I), RJ11 and RJ45 connections for anemometer data and CAT-5 connectivity (J, K), microphone assembly (L), weather tight AC power input (M), and solar sensor assembly for light  
15 intensity measurement (N). 3D-printed parts include the gas sampling manifold, rain hoods, exhaust and microphone mounting bracket, solar sensor interface, and PCB mounting scaffold. As described in the text, the gas and particle sampling inlets are decoupled, with the OPC-N2 sampling through the bottom-face of the enclosure to protect from liquid water penetration.

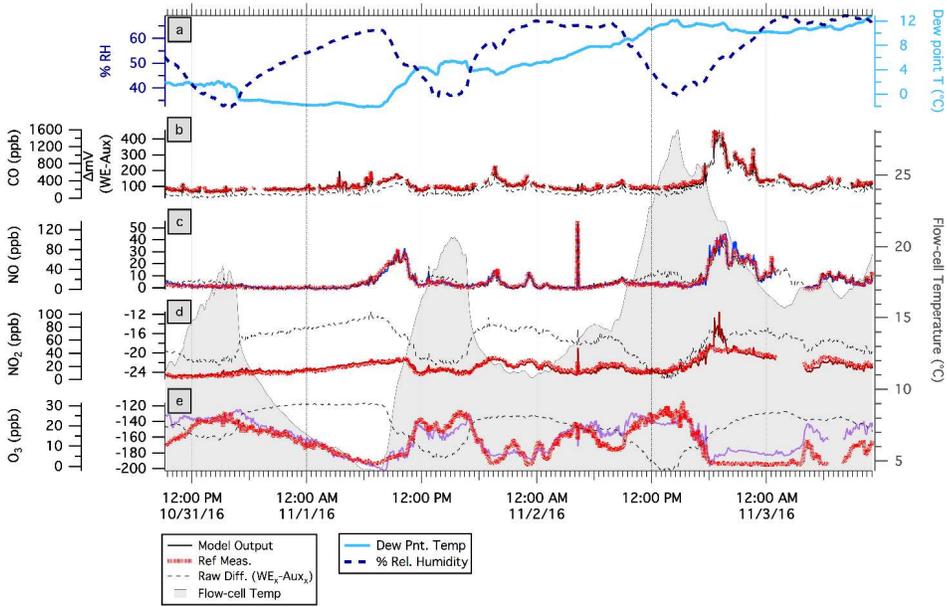


10

**Figure 2.** Map showing the locations of the two ARISense systems (yellow circles) and the four metro-Boston DEP monitoring stations (each marked with a red circle). The two ARISense systems were co-located with reference stations at the Harrison Avenue site (in Dudley Square, Roxbury) and Von Hillern Ave. site (~35 feet from I-93 North) in Dorchester. The data presented in this manuscript were obtained from the Dudley Square location, an urban neighborhood site, primarily impacted by local combustion sources and vehicles operating on secondary routes in close proximity to the area. Green markers are shown to indicate the positions of 4 additional sensor nodes deployed as part of the Dorchester Air Quality Sensor System (DAQSS) project, which pre-dated the development of ARISense.

15

20

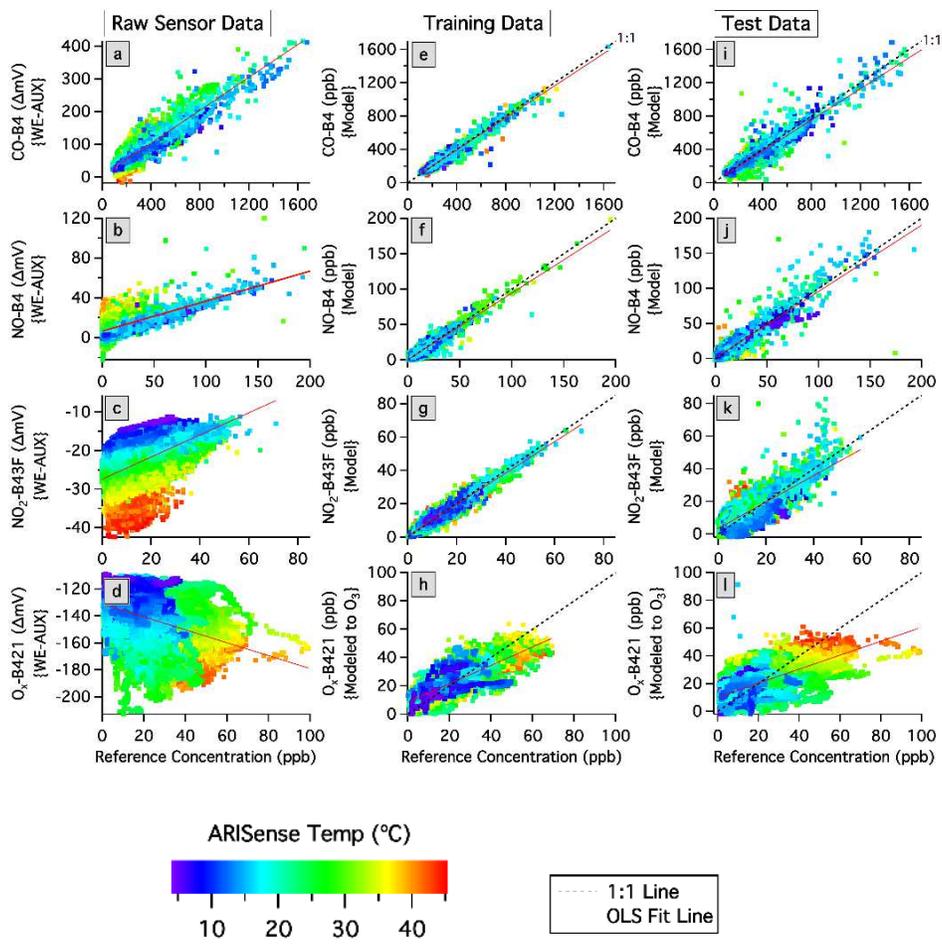


**Figure 3.** (a) Relative humidity (solid blue line) and dew point temperature (dashed blue line), (b-e) raw differential sensor output (dashed line), reference measurement (thick red line) and model output (thin line) for a 3 day period during the test part of the 4.5 month deployment. Temperature is indicated with the grey shaded area.

10

15

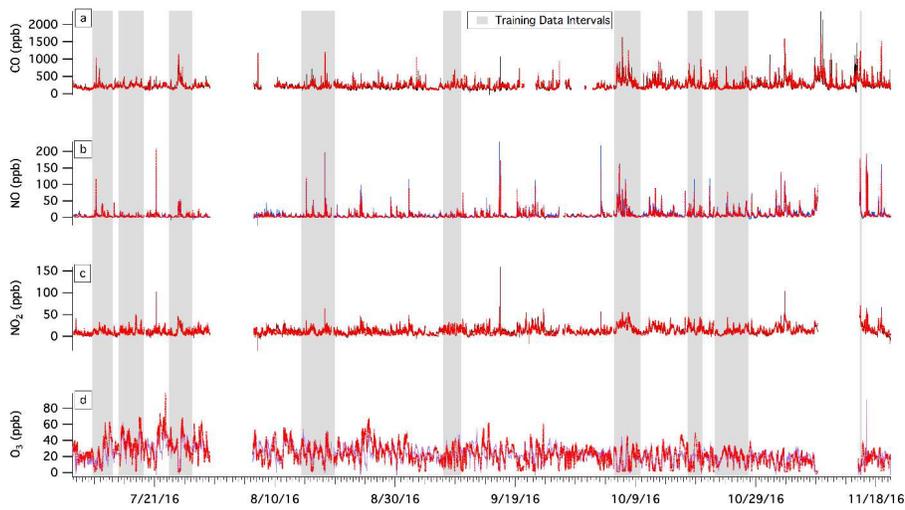
20



5 **Figure 4.** Correlation plots for all electrochemical sensors versus reference measurements for (a-d) raw sensor differential voltage signals, (e-h) model output concentrations for the training data (~25% of data), and (i-l) model output concentrations for test data (remaining ~75% of data). All data shown are 5-minute average values with each data point colored by flow-cell temperature. The linear regression fit-line (solid red line) is shown in all panels, and a  
 10 1:1 line (dashed black line) is shown in panels e-l.

**Formatted:** FigCaption, Line spacing: single, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

**Deleted:** ¶  
 ¶  
 ¶  
 ¶  
 ¶



5  
**Figure 5.** Time-series of 5-min averages of the model output (sensor) and reference gas concentrations. Grey shaded areas indicate time periods over which the model was trained. A unique set of input parameters was used to train the HDMR model for each of the different electrochemical cells. Approximately 25% of the data was used for training and  
 10 the remaining 75% was used to test the models.

In order to demonstrate the varying effect of temperature on the different electrodes, Figure 5 shows raw 2-min average output currents for both the working (darker hue) and auxiliary (lighter hue) electrodes for each electrochemical sensor over a 48-hour time interval. Reference concentrations of each pollutant species are plotted (as red dashed-lines) on the right-hand axes for comparison, and the flow-cell temperature is displayed as a filled histogram in the background of the time-series. Temperature changes in excess of 22 C are observed in as little as 10 hours. Offsets are observed between the working and auxiliary electrodes of all four sensors, but the extent to which the working/auxiliary pair track one another as environmental conditions change is highly sensor-dependent. The CO-B4 sensor appears to be fairly insensitive to the temperature changes encountered, with a positive-going differential reflecting changes in the ambient concentration of CO. In contrast to the muted temperature response of the CO sensor, the NO-B4 sensor shows a strong temperature dependence in both the working and auxiliary signal, with some evidence of lagging temperature response in the auxiliary channel at the highest temperatures encountered. The NO<sub>2</sub>-B43F shows a clear temperature dependence in the working electrode, but the auxiliary electrode shows less of a response. However, in a few instances in the 48-hour window displayed in the figure, it appears that the NO<sub>2</sub>-B43F auxiliary response to rapid decreases in temperature (coincident with sunset) is opposite the trend observed for the working electrode. The Ox-B421 sensor exhibits a temperature-dependent response that largely tracks between the auxiliary and working electrodes.

The next step in the analysis was to determine whether statistically significant correlations between raw EC-sensor outputs and reference pollutant concentrations could be extracted from the data shown in Fig. 3 plus the environmental data in Fig. S1.

(remainder of the 4.5 month deployment, excluding the training data)

by systematically testing all possible input parameter combinations and then evaluating the resultant model's performance against the full co-location dataset. An example of two different input matrices for determining NO<sub>2</sub> concentrations is provided in the Supplemental Material (see Fig. S2).

different electrochemical sensors, and the working WE and auxiliary AUX electrodes

Subscript

Subscript

plots for the full dataset analyzed with the HDMR models trained on the (35%) subset of ambient data. Data shown are

- Page 10: [9] Deleted** Leah Williams 8/20/2017 4:11:00 PM  
 Correlation plots of sensor-derived and reference concentrations for the entire dataset are shown in Fig. 4i-l.
- Page 10: [10] Formatted** Leah Williams 8/20/2017 4:31:00 PM  
 Subscript
- Page 10: [11] Formatted** Leah Williams 8/20/2017 4:18:00 PM  
 Font: Italic
- Page 10: [12] Formatted** Leah Williams 8/20/2017 4:18:00 PM  
 Superscript
- Page 10: [13] Formatted** Leah Williams 8/20/2017 4:19:00 PM  
 Font: Italic
- Page 10: [14] Formatted** Leah Williams 8/20/2017 4:19:00 PM  
 Superscript
- Page 10: [15] Formatted** Greg Magoon 8/21/2017 3:30:00 PM  
 Font: Italic
- Page 10: [16] Formatted** Greg Magoon 8/21/2017 3:30:00 PM  
 Font: Italic
- Page 15: [17] Deleted** Eben Cross 8/22/2017 1:21:00 PM

**Table 4.** [EC1]Comparisons to published results utilizing Alphasense electrochemical sensors

Studies	Temporal Resolution (min)	N <sub>data pts</sub>	Slope	r <sup>2</sup>	RMSE E (ppb)	MA (ppb)	MBE (ppb)	f <sub>train</sub> * (%)
<b>CO-B4 SENSOR</b>								
Jiao et al., 2016	60	2640-2664	7.99E-4-8.09E-4	0.63-0.68	NR	NR	NR	
Castell et al., 2017	15	6912		0.36	170.99			
Zimmerman et al., 2017	15							
This Work	5							
<b>NO-B4 SENSOR</b>								
Jiao et al., 2016	60	2640-2664	0.883-0.892	0.77-0.87	NR	NR	NR	
Castell et al., 2017	15	6912		0.74	16.35			
This Work	5							
<b>NO2 SENSOR</b>								
Jiao et al., 2016	60							
Castell et al., 2017	15							
Zimmerman et al., 2017	15							

This Work 5

---

Ox SENSOR

Jiao et al., 2016 60

Castell et al., 2017 15

Zimmerman et al., 2017 15

This Work 5

---

NR = not reported in manuscript