Atmospheric
Measurement
Techniques

Supplement of

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

Eben S. Cross et al.

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

Supplemental Material.

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	$r^2 = 1 - \dfrac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	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	$RMSE = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$	Standard deviation of difference between model output and reference values. Measure of accuracy. Sensitive to outliers.
Mean absolute error	MAE	$MAE = \dfrac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$	Average of the absolute error. Disregards the direction of under- or over-prediction.
Mean bias error	MBE	$MBE = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$	Average error. Indicates if model output values are biased high or low relative to reference values.

5

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).

10 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.

15 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

20 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).

25 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.

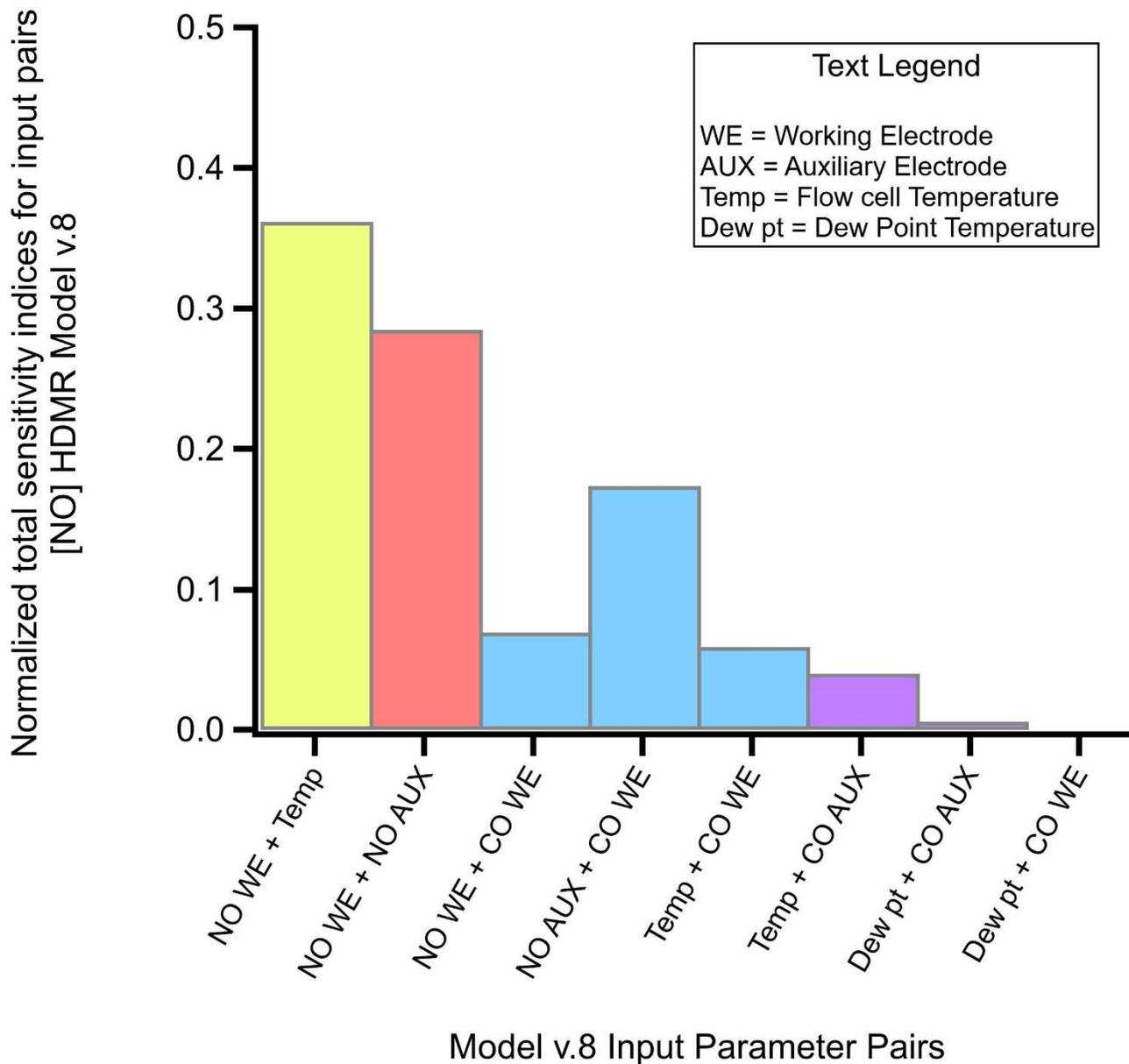| Model v.	8	9	10	11	12	13
CO AUX	x				x	x
CO WE	x			x	x	x
NO AUX	x	x			x	x
NO WE	x	x	x	x	x	x
NO2 AUX					x	x
NO2 WE					x	x
Ox AUX					x	
Ox WE					x	
Dew point	x	x	x	x	x	x
Temperature	x	x	x	x	x	x
CO2 (voltage)					x	x
RMSE (ppb) { test }	3.38 { 4.52 }	5.09 { 5.86 }	6.75 { 7.16 }	4.88 { 5.53 }	2.58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

5

Figure S1. Normalized total sensitivity indices of each significant (contribution > 0.001) 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.

5

10
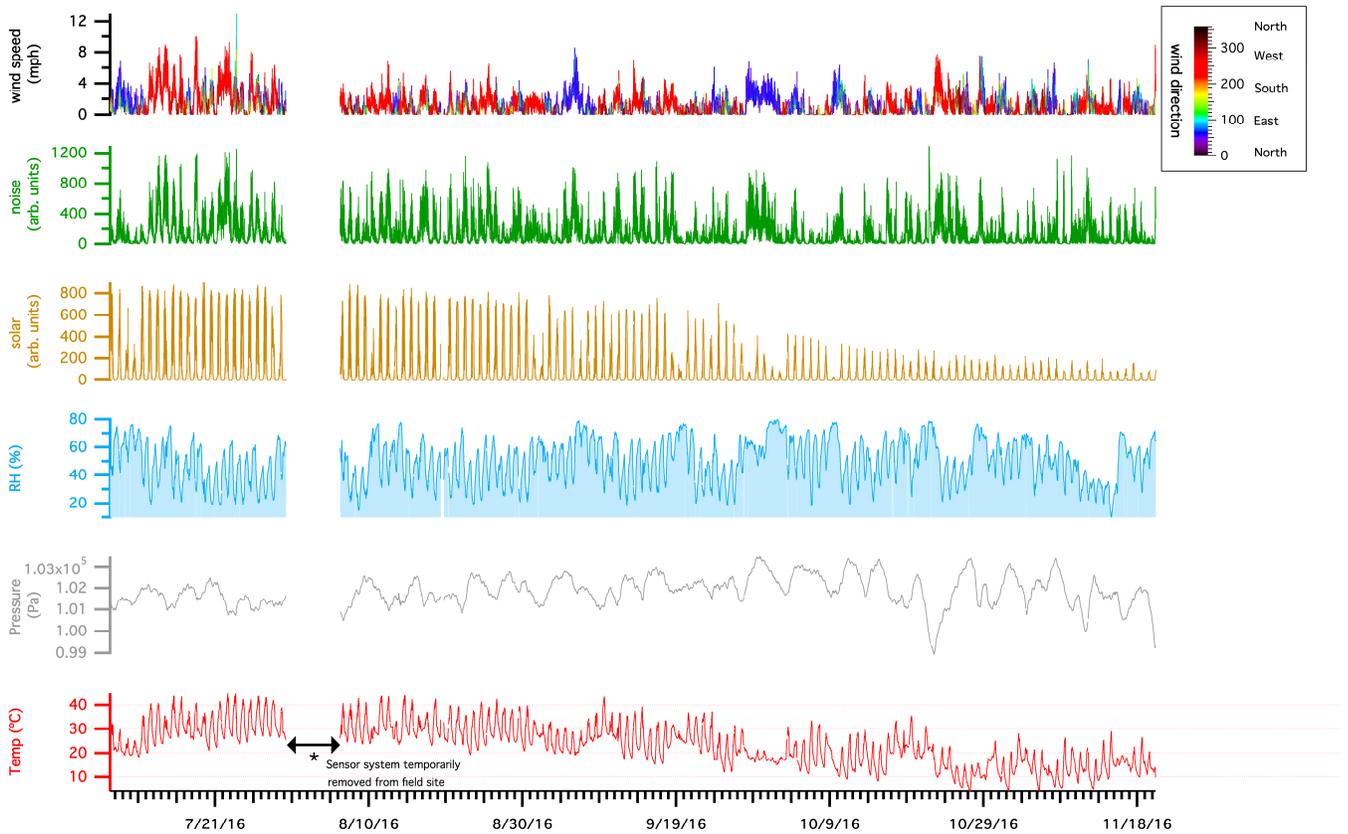
Figure S2. Continuous 5-min average non-pollutant data acquired with the ARISense system, tracking ambient variability in temperature, pressure, humidity, solar intensity, ambient noise, and wind speed and direction at the Roxbury DEP monitoring site. Meteorological conditions cover a transition from warmer (summer) to cooler (late-fall) seasons in the NE US. Temperature and humidity measurements shown reflect the conditions within the gas-sampling flow cell of the integrated system, the metrics most relevant to modelling EC interference effects derived from environmental conditions. The break in ARISense data from July 30 – August 6 was intentional, corresponding with temporary removal of the system from the site to execute a short-term experiment at another location. Short term interruptions in sampling were primarily caused by power interruptions at the site. The ARISense system is programmed to automatically initiate sampling on power-up, and was able to regain stable operation each time power was restored at the site. One data-feature of note is the high degree of correlation between the ambient noise level and wind speed. For ARISense v1.0 systems, the microphone integration lacked a wind-screen, resulting in wind-derived noise dominating the signal for conditions in which wind speeds ≥ 4 MPH. Subsequent versions of the system will include a wind-screen to minimize wind-derived signal in the microphone and improve audible resolution of other noise sources in close proximity to the node.
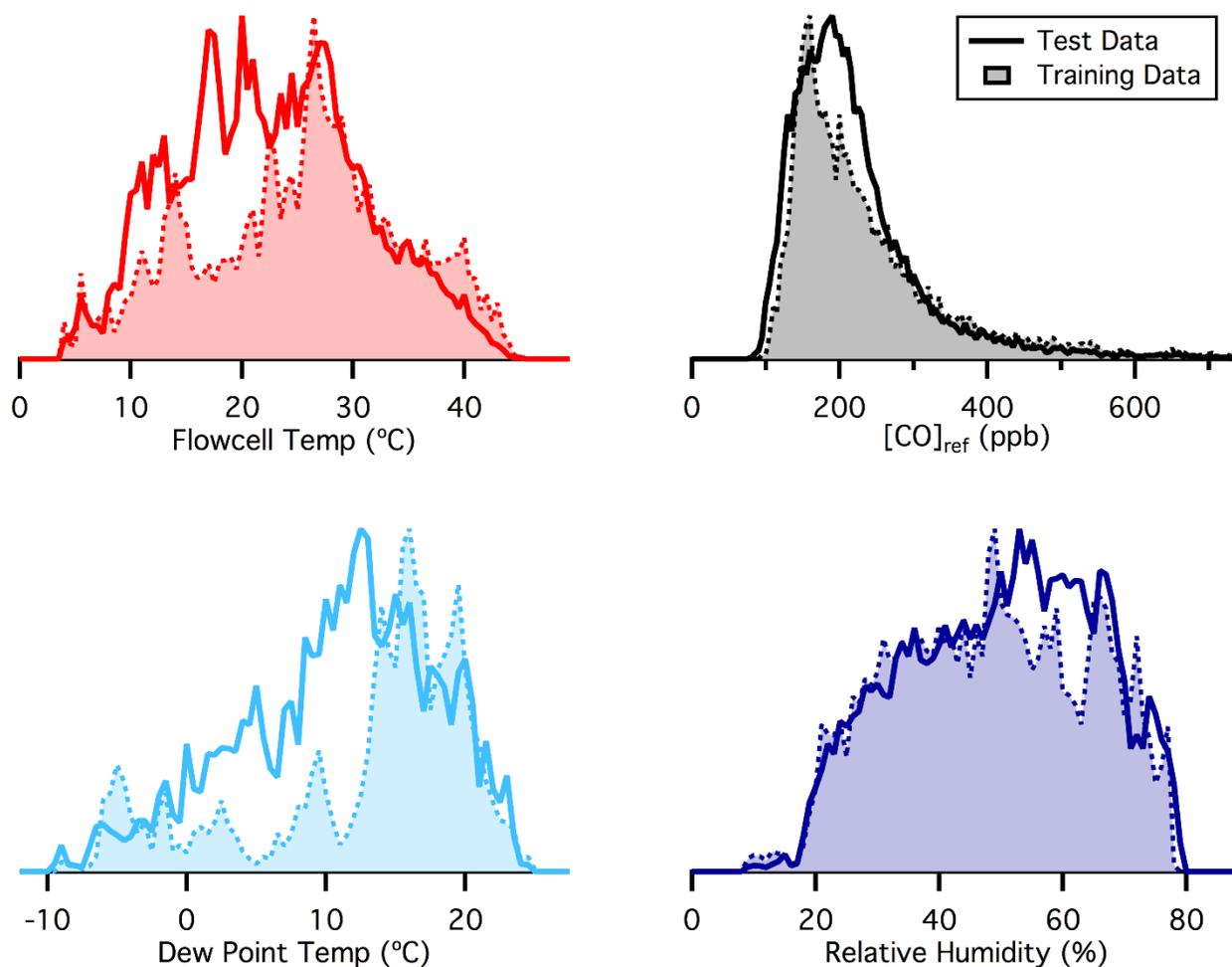
Figure S3. Distributions of temperature, reference measurement, dew point temperature, and relative humidity for the training data for the CO HDMR model (dashed lines/shaded) and the test (solid lines) data. In the case of CO, the training data distributions were generated from 27% of the total available 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.