Characterizing the Aging of Alphasense NO2 Sensors in Long-term Field Deployments

Jiayu Li1,2, Aliaksei Hauryliuk1, Carl Malings1,3,4,5, S. Rose Eilenberg1, R. Subramanian1,3,4, Albert A. Presto1,\*

1Center for Atmospheric Particle Studies (CAPS)

Carnegie Mellon University

Pittsburgh, PA 15213

2Air and Aerosol Sensing Group (AASG)

University of Minnesota, Twin Cities

Minnesota, MN 55108

3OSU-EFLUVE - Observatoire Sciences de l’Univers-Enveloppes Fluides de la

Ville à l’Exobiologie, Université Paris-Est-Créteil, CNRS UMS 3563

Ecole Nationale des Ponts et Chaussés, Université de Paris, France

4Laboratoire Interuniversitaire des Systèmes Atmosphériques, UMR 7583, CNRS

Université Paris-Est-Créteil, Université de Paris, Institut Pierre Simon Laplace

Créteil, France

5NASA Postdoctoral Program Fellow

Goddard Space Flight Center

Greenbelt, MD 20771

\*Corresponding author: Albert A. Presto

Email: [apresto@andrew.cmu.edu](mailto:apresto@andrew.cmu.edu)

Phone: 412-721-5203

**Abstract**

Low-cost NO2 sensors have been widely deployed for atmospheric sampling. While their initial performance has been characterized, few studies have examined their long-term degradation. This study focused on the performance of Alphasense low-cost NO2 sensors (NO2-B42F and NO2-B43F) over four years (2016 – 2020). A total of 29 NO2 sensors from ten batches were collocated 78 times at two sites with reference instruments. Raw signals from “functional” NO2 sensors correlated linearly with reference NO2 concentrations. After long-term deployment, sensor raw signals started to deviate from reference NO2 concentrations due to sensor aging, an accumulated effect after sensor unpacking. Several sensors eventually became “non-functional” as sensor raw signals showed no correlation with reference NO2 concentrations. Sensor aging and non-functionality may be primarily caused by expiration of the ozone (O3) scrubber built into these sensors, so that sensors responded to both ambient NO2 and O3. The influence of O3 on sensor response is quantified through the permutation importance method. Most of the sensors are non-functional after approximately 200-400 days of deployment, and no sensor was functional after 400 days of deployment. This result agrees well with the estimated lifetime of the built-in ozone scrubbers considering the ambient ozone concentration in the Pittsburgh area where these sensors were deployed. To ensure reliable data quality in long-term field deployments, we recommend collocating NO2 sensors with reference instruments regularly after 200-400 days of deployment to identify and replace non-functional sensors in a timely manner.

**Keywords**: Low-cost NO2 sensor, long-term deployment, sensor aging, ozone scrubber, malfunction identification.

Nitrogen dioxide (NO2) is an air pollutant that can cause respiratory and cardiovascular diseases at low concentrations and may cause death after exposure to sufficiently high concentrations.1 Ambient NO2 concentration generated from traffic is correlated with a higher incidence of asthma, wheezing bronchitis, ischemic stroke, and myocardial infarction.2 Due to these adverse health impacts, the U.S. Environmental Protection Agency’s (EPA’s) National Ambient Air Quality Standards (NAAQS) specify the maximum ambient NO2 concentration (annual average less than 53 ppb, 1-hour average not to exceed 100 ppb) and mandate routine monitoring.

The federal reference method (FRM) for NO2 sampling requires expensive monitoring equipment. This limits the number of sampling locations and leads to low spatial resolution. To improve spatiotemporal resolution with low cost and maintenance requirements, electrochemical NO2 sensors, whose price is approximately 1-2% of the regulatory methods, have been studied extensively as an alternative.3 After characterization and calibration, these NO2 sensors have been deployed with high spatiotemporal resolution for air quality measurement, and the possibility of using these sensors for exposure estimation has been discussed.3a, 4

The Alphasense NO2 sensor is one of the most widely-used original equipment manufacturer (OEM) air quality sensors.3a, 3b, 3f-i, 5 It is an amperometric sensor that detects NO2 on a solid sensing electrode immersed between the sampled gas and an internal liquid electrolyte.3g, 6 When NO2 molecules react with the electrolyte, a current is generated that can be correlated with the NO2 concentration.6 Linear regression, multivariate linear regression, and machine learning algorithms have been used in previous literature to calibrate these sensors,3b, 3h, 3i, 7 with R2 values for model validation ranging from 0.27 to 0.95.3b, 3e, 3f, 3i, 7b, 8 Environmental conditions, such as temperature and relative humidity, are usually included in the calibration algorithm to improve the data quality.3b, 4a, 9 Some calibration algorithms, especially machine learning algorithms,3b, 3h, 3i also incorporate concentrations of other measured species.

The ozone (O3) concentration is a concern among these additional parameters since NO2 and O3 are reducible at similar potentials at carbon and gold electrodes.8b, 10 The sensor responding to untargeted O3 raises concerns of cross-sensitivity. Ideally, sensors should respond to the targeted gas (NO2) only, as interference from an untargeted gas may critically bias their accuracy. Alphasense NO2 sensors have an O3 scrubber at the inlet to eliminate the interference from ambient O3. The O3 scrubber lifetime reported by the manufacturer is 500 hours when exposed to 0.5 ppm O3 (250 ppm-hr of O3 exposure). The 2019 annual average O3 concentration was 25.2 ppb at an urban reference site in Pittsburgh, PA, USA, where this study was conducted. Therefore, the expected lifetime of the O3 filter is approximately 413 days or ~14 months.

Although the calibration and application of Alphasense NO2 sensors have been studied extensively, few studies have examined their long-term stability. Long-term stability is essential for maintaining reliable data quality. Malings et al. (2019) evaluated the performance of Alphasense NO2 sensors over 18 months in Pittsburgh. Tested NO2 sensors showed a consistent performance at collocation sites with r values ranging from 0.5 to 1. No significant seasonal trend in sensor performance was observed. In this study, we tracked the performance of multiple NO2 sensors over a four-year period (2016 to 2020). Sensors were unpacked and deployed gradually over four years; the longest deployment of an individual sensor was approximately 1,200 days. Sensors were periodically brought back to reference-grade monitoring sites for collocation with reference instruments. These periodic collocation data sets enable us to evaluate the sensor performance over a long time. We also investigate the influence of temperature, relative humidity, and ozone concentration on sensor aging.

Methods

RAMP monitor and NO2 sensors

The Real-time Affordable Multi-Pollutant (RAMP) monitors are end devices of a low-cost air quality monitoring system collaboratively designed by SenSevere (now part of Sensit Technologies) and the Center for Atmospheric Particle Studies (CAPS) at Carnegie Mellon University (CMU). Each monitor has an optical particulate matter (PM) sensor and multiple electrochemical gas sensors to measure carbon monoxide (CO), O3, sulfur dioxide (SO2), and NO2. The NO2 concentration is measured by Alphasense NO2-B42F and NO2-B43F (Essex, United Kingdom) four-electrode electrochemical sensors. Calibration algorithms, evaluation procedures, and applications of RAMPs have been elaborated in previous publications.3b, 3h, 11 Before initial deployment, RAMPs were calibrated by collocating with reference instruments. During the deployment, RAMPs were brought back for collocation periodically to ensure data quality. The collocation datasets for calibrating NO2 sensors are mainly collected from two reference monitoring sites described in detail below.

From 2016-2020, a total of 102 Alphasense NO2 sensors from ten batches (batch numbers 19, 29, 47, 48, 55, 66, 68, 69, 106, 108) were deployed on 28 RAMP monitors. A unique serial number and a shared batch number were used to identify and track each NO2 sensor. Because sensors were replaced when apparent failures occurred during collocations, the number of sensors is higher than the number of RAMP monitors. The characteristics and performance of sensors from the same batch are supposed to be similar. Batches 47, 55, 66, and 106 were purchased and deployed between 2016 and 2017, while batches 19, 48, 49, 68, 69, and 108 were purchased and deployed between 2018 and 2019.

Sensor collocation calibration

RAMPs were repeatedly collocated at the “Lawrenceville” and “CMU campus” monitoring sites for field calibration (Figure S1). The reference instruments at these two monitoring sites provide ground truth NO2 concentrations to evaluate the performance of NO2 sensors. The “Lawrenceville” site is a monitoring station operated by the Allegheny County Health Department (ACHD) and is part of the National Core Multipollutant Network (NCore, Air Quality System (AQS) I.D. 42-003-0008). ACHD Lawrenceville is a population-based community-oriented monitoring site in a neighborhood with urban residential and commercial land use characteristics. This site uses Teledyne Advanced Pollution Instrumentation (TAPI) 200EU (detection limit: 50 ppt) to measure NO2 concentration.

The “CMU campus” site is located on the Carnegie Mellon University campus (Figure S1). The CMU campus site used a 2B Technologies Model 405nm NO2 monitor (detection limit < 1ppb). RAMP monitors were brought to the CMU campus site in 2016 and 2017 for collocation calibration. Unlike ACHD Lawrenceville, the CMU campus site was not operated continuously throughout the year. Sensors were collocated at CMU from May – November 2017. More information regarding the location and NO2 concentration range of these two reference sites is included in the supplemental information.

Most sensors were collocated either at Lawrenceville or the CMU campus site before field deployment (e.g., immediately or shortly after the sensor was unpacked from the package). Collocations were not planned on a fixed schedule but were done any time one of the gas sensors required replacement or any other maintenance issues occurred (e.g., intermittent communication with the data server). A total of 78 collocation records from 29 NO2 sensors were identified. Collocations happened between 0 and 1299 days (~3.5 years) after the unpacking of the NO2 sensors, enabling us to track the performance of sensors over a long time. The length of each collocation period varied. The shortest collocation durations at the CMU campus site and Lawrenceville were four days and one day. The longest collocation durations at the CMU campus site and Lawrenceville were 111 and 118 days respectively. In each collocation, the NO2 concentration, O3 concentration, temperature, and relative humidity from each site were paired with raw NO2 sensor signals for aging analysis.

Kendall rank correlation

Collocation datasets usually have many more data points in the low concentration range, so we pre-treated datasets to adjust the weight of data points in the low and high concentration ranges. To eliminate the interference of a large amount of data in the low concentration range, we binned the original data into 1 ppb increments based on the reference NO2 concentrations. Another advantage of binning data is that it makes up the data size difference among collocations. The original size of collocation datasets varied dramatically depending on collocation durations. After binning the data, the size of each collocation dataset is approximately the same.

Datasets were analyzed with Kendall rank correlation to measure the ordinal association between sensor raw signals and reference NO2 concentrations. The Kendall rank correlation (Eq. (1)) is a robust non-parametric test without assuming any data distribution.12 It outperforms Spearman rank correlation in handling data outliers. It can also manage the no slope or infinite slope situation when there is no relation between sensor raw signals and reference NO2 concentrations, which is problematic in Pearson correlation. Kendall’s coefficients are between 0 and 1, and larger values represent a strong association.

In Eq. (1), the denominator is the total number of pairings of *n* samples,and *nc* and *nd* are numbers of concordant and discordant, respectively. Data need to be placed in order (as shown in Eq. (2)) and be paired to calculate numbers of concordant and discordant. In Eq. (2), the *S* and *R* represent sensor raw signals and reference NO2 concentrations, respectively. For a pair of data, if the sign of (*Si* – *Sj*) is the same as the sign of (*Ri* – *Rj*), it is a concordant pair. Otherwise, it is a discordant pair.

(1)

, where (2)

Results and discussion

Sensor performance over time

Figure 1 shows the aging evolution of a single NO2 sensor over three years. This NO2 sensor was from batch 66 and was installed on RAMP 111. After unpacking in August 2016, it was brought to collocation sites several times until June of 2019. Panels (a-b), (c-d), (e-f), and (g-h) are results of collocations beginning at 80, 135, 321, and 961 days after unpacking, respectively. The duration of the collocation was 5, 16, 74, and 55 days, respectively. Black dots in the left panels show the correlation between the sensor raw signals (mV, here defined as auxiliary electrode voltage minus working electrode voltage) and reference NO2 concentrations. Raw signals are linearly correlated with reference NO2 concentrations in collocations that happened 80 days and 135 days after unpacking, indicating the sensor works as expected or is “functional.” A smaller raw sensor signal (i.e., a more negative value) corresponds to a higher NO2 concentration.

The data in Figure 1(c) are shifted to the left compared to the data in Figure 1(a) that are outlined by the orange dashed line. We refer to this shift as sensor drift. Raw signals between -20 to -15 mV correspond to NO2 concentrations higher than 30 ppb in Figure 1(a), while in Figure 1(c), the same sensor outputs correspond to reference NO2 measurements ranging from 2.3 to 36.2 ppb. This drift issue would lead to occasional overestimation in the low concentration range if the previous calibration model was used to convert the raw signals to concentrations.

Figures 1(e) and (g) show collocations that occurred 321 and 961 days after unpacking. Sensor raw signals show almost no relationship with NO2 concentrations at these collocations. In Figures 1(e) and (g), when the NO2 concentration was below 10 ppb, the sensor’s raw signals ranged from -30 to 20 mV, basically the full scale of the sensor response. Such a compromised performance made it challenging to invert raw signals to reference NO2 concentrations even with advanced algorithms. Based on such performance, we consider this sensor to be a non-functional sensor, not suitable for further deployment.

Panels on the right of Figure 1 show the correlation and difference between the sensor’s “model inverted” NO2 concentrations and reference NO2 concentrations. To convert sensor raw signals to NO2 concentrations, we used 80% of the 80-days collocation dataset to train a hybrid (HY) model as described by Malings et al. (2019) and the remaining 20% of the dataset for validation. The HY model uses a polynomial regression for the highest 30% of NO2 concentrations and a RF model for the rest of the data (lower concentrations). The relation between the HY model inverted NO2 concentrations and reference NO2 concentrations are shown as brown dots in the right panels. Bars in the right panels are absolute errors between predictions and reference NO2 concentrations (HY NO2 concentration – reference NO2 concentration) grouped by one ppb reference NO2 concentration increments. If the HY inverted NO2 concentrations are close to actual NO2 concentrations, shorter bars can be observed, showing smaller absolute errors.

In Figure 1(b), brown dots align well with the 1:1 ratio (black dashed line), indicating the NO2 concentrations inverted from HY model are close to reference NO2 concentrations. Short pink bars in Figure 1(b) also indicate a good agreement between HY model and reference NO2 concentrations. With sensor aging, in Figures 1(d), (f), and (h), absolute errors become larger and brown dots deviate from the black dashed line. The HY model overestimates NO2 concentrations in the low concentration range and underestimates NO2 concentrations in the high concentration range. Apart from raw signals from NO2 sensors, the HY model also includes the response of other air quality sensors in the RAMP and environmental conditions (RH and T) as variables,3b, 3h which potentially weakens the impact of sensor aging. Although the correlation between raw sensor signals and reference NO2 concentrations is very weak in Figures 1(e) and (g) (R2 = 0.03 and 0.02), the HY inverted NO2 concentrations are still roughly linearly correlated with reference NO2 concentrations in Figures 1(f) and (h) (R2 = 0.36 and 0.53). This is because the RF model includes responses from multiple sensors (e.g., O3, CO, RH, and T sensors) as input and thus dilutes the impact of NO2 sensors. Even if raw sensor signals correlate weakly with the ambient NO2 concentration, the model can still rely on other sensors to predict reasonable NO2 concentrations.

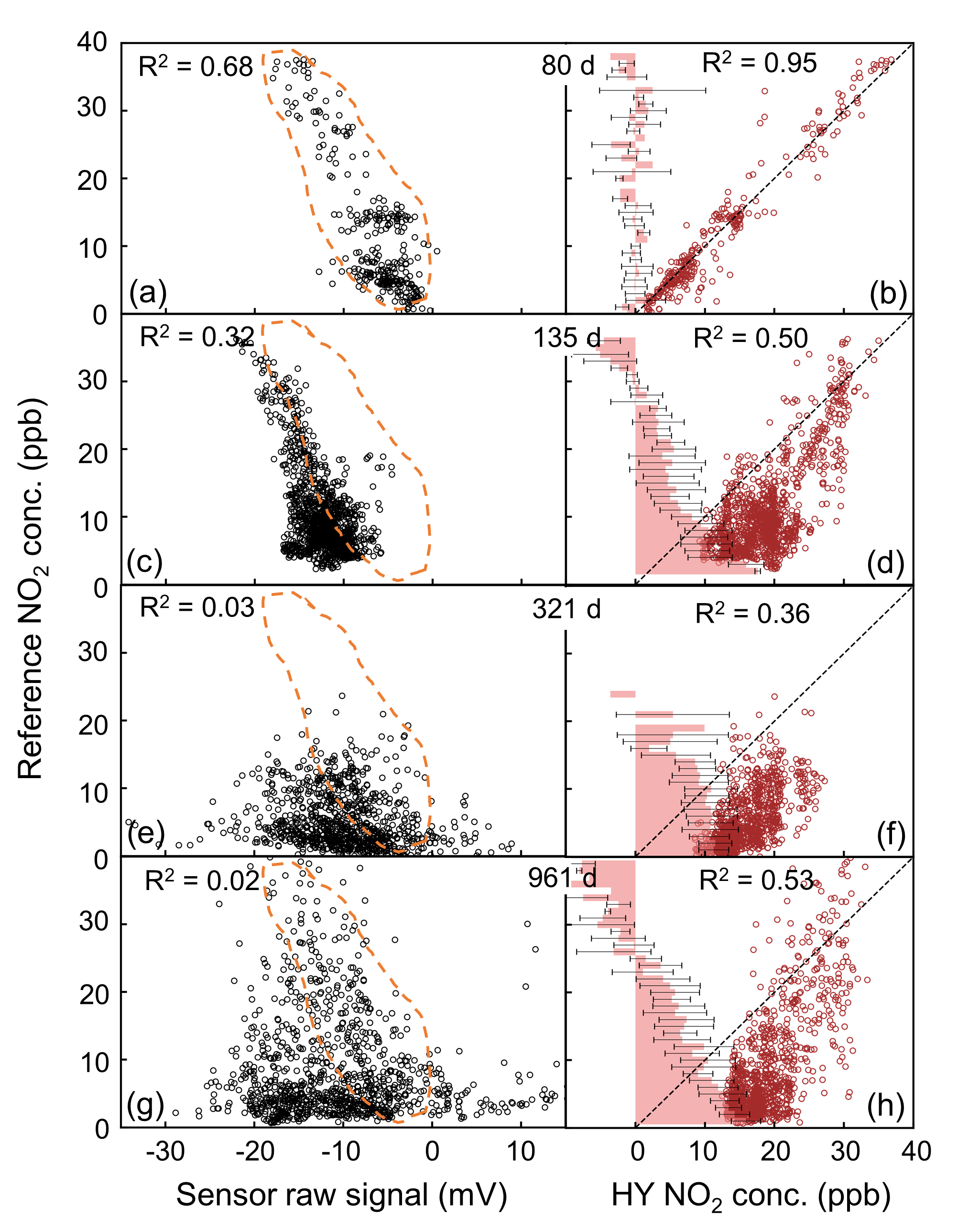


Figure 1. The aging of RAMP 111 (batch 66) NO2 sensor in three years. Subplots (a-b), (c-d), (e-f), and (g-h) are collocation results beginning from 80, 135, 321, and 961 days after unpacking, respectively. The duration of the collocation was 5, 16,74, and 55 days, respectively. Subplots (a), (c), (e), and (g) demonstrate that the correlation between sensor raw signals and reference NO2 concentrations degrade over deployment time. The orange dashed line shows the extent of the data from the collocation at 80 days. Subplots (b), (d), (f), and (h) show the correlation between the hybrid (HY) model inverted NO2 concentrations and reference NO2 concentrations. The bar plots in subplots (b), (d), (f), and (h) are errors between HY NO2 concentrations and reference NO2 concentrations grouped by one ppb increments.

Sensor performance evaluation

Figure 2 displays Kendall’s coefficients of collocation datasets over deployment time. The x-axis, labeled as “days after unpacking,” is the number of days from sensor unpacking to the start of the collocation, and the y-axis shows Kendall’s coefficients. Different markers represent sensors of different batches, and the color of the marker represents the maximum NO2 concentration during the collocation. With sensor aging, τ values decrease over time. The majority of data points with higher τ values (>0.6) are clustered to the left side, showing that most sensors started as functional sensors with raw signals strongly dependent on NO2 concentrations. One exception is a sensor from batch 48 that demonstrated a low τ value (<0.2) on the first day of deployment; however, the NO2 concentration during this collocation is relatively low (below 25 ppb), which could cause such a result. This heterogeneity demonstrates that there could be variations in sensor performance, even for new sensors.

The range of NO2 concentration can influence Kendall’s coefficients during collocations. The color of the markers shows the maximum NO2 concentration during the collocation. Both a functional and a non-functional sensor could yield low τ values if they were exposed to low NO2 concentrations throughout the collocation. Thus, a low τ value may not be the absolute standard to determine the status of an individual sensor. On the other hand, a low τ value under a high NO2 concentration is persuasive evidence of a non-functional sensor (e.g., yellow symbols after 600 days). At the same time, a high τ value under a low NO2 concentration indicates a functional sensor, shown as blue and green markers before 200 days.

Noticeable sensor performance degradation happened approximately 200-400 days after deployment. Before 200 days, most sensors demonstrated high τ values regardless of the maximum NO2 concentration. After approximately 200 to 300 days in the field, some sensors started to show low τ values in collocations. After 400 days, almost all sensors showed lower performance (τ <0.5) and can be considered non-functional. This agrees well with the approximate lifetime of the O3 scrubber calculated based on the sensor datasheet and the ambient O3 concentration in the Pittsburgh area, well short of the expected 2-year lifetime specified by Alphasense.

There are some yellow markers after 400 days, showing that some sensors were collocated under high NO2 concentrations, during which they still yielded a similar low τ value compared to sensors exposed to low NO2 concentrations. This indicates that sensors were non-functional even at high NO2 concentrations; the weak correlation between sensor raw signals and reference NO2 concentrations was not influenced by the ambient NO2 concentration. Figure 2 also indicates that sensor aging is irreversible since τ values remain low after long-term deployment. There are no instances of a sensor moving from a non-functional state back to a functional state.

Chart, scatter chart

Description automatically generated

Figure 2. Kendall’s coefficients (τ) change with days after unpacking. Days after unpacking refers to the duration from unpacking to the start of the collocation. Different markers represent sensors from different batches. The color of each dot represents the maximum NO2 concentration sensors were exposed to during the collocation. Kendall’s coefficients decreasing over time is an important sign of sensor aging.

Parameters influencing aging

Previous sections demonstrated that sensor performance degraded over time. This section evaluates the influence of T, RH, and ozone concentration on sensor aging. It is difficult to fully decouple interferences from T, RH, and O3 because these three parameters are all coupled through atmospheric chemistry. Usually, T and RH are anti-correlated, and T and O3 are generally positively correlated. Examples of both functional and non-functional sensors are analyzed below to identify each factor’s influence.

Figures 3(a), (c), (e) show the performance of a functional sensor, whose raw signals demonstrated good linearity against NO2 concentrations. The points in each panel are colored by T, RH, or O3 concentration. These confounders may magnify the noise level, especially in the low concentration range. Some non-linear behavior can be observed during periods of high temperature and high ozone. At low NO2 concentrations (< 5 ppb), sensor signals are variable (0 – 30 mV). This may be due to sensor cross-sensitivity to either the high O3 concentration or high temperature. Since NO2 and O3 are reducible at similar potentials, NO2 sensors should respond similarly to NO2 and O3. For a functional NO2 sensor, a smaller signal (more negative) would be generated with higher NO2 concentrations. Likewise, a smaller or more negative signal should be expected under high O3 concentrations. However, shown as yellow dots in Figure 3(e), NO2 sensors showed more positive signals under high O3 concentrations, which excludes the high O3 concentration as the cause of the cross-sensitivity issue. Hence, for functional sensors, we suspect that high temperatures cause the more positive sensor raw signals at low NO2 in Figure 3.

Chart, scatter chart

Description automatically generated

Figure 3. The performance of functional and non-functional sensors varying with temperature, relative humidity, and O3 concentration. Figures 3(a), (c), and (e) are results from RAMP 120 with a functional batch 48 NO2 sensor collocated in Lawrenceville from May 17th to June 28th, 2018, 7 days after unpacking. Figures 3(b), (d), and (f) are results for RAMP122 with a batch 66 NO2 non-functional sensor collocated in Lawrenceville from April 2nd to May 7th, 2018, 601 days after unpacking.

We repeated the same analysis for non-functional sensors, whose raw signals demonstrated a very weak or no response to NO2 concentrations (τ < 0.3). Figures 3(b), (d), and (f) are April to May 35-day collocation results, 601 days after a batch 66 sensor was installed on RAMP 122. Due to the non-linear response of non-functional sensors, it is even more challenging to identify each confounder’s influence. Relative humidity does not significantly influence the performance of non-functional sensors, as shown in Figure 3(d). Still, with high temperature and high O­3 concentration, the bright dots in Figures 3(b) and (f) lead to large raw signal variation in the low NO­­2 concentration range.

Comparing sensor performance in the low NO2 concentration range from functional and aged sensors, a very obvious difference is that a functional sensor seldom generated small (more negative) signals under low NO2 concentrations, while a non-functional sensor frequently reported small (more negative) signals regardless of actual NO2 concentrations. These small signals from non-functional sensors would lead to significantly overestimated NO2 concentrations resulting in the triangular data distributions in Figure 3(b), (d), and (f). The signals located in each panel’s lower-left corner are usually generated when the O3 concentration was high. As we discussed in the previous section for functional/newer sensors, high T tends to generate larger (more positive) signals. High O3 concentrations should react on the sensor similarly to NO2 and generate smaller (more negative) signals. Thus, O3 is likely causing false small (more negative) signals on non-functional sensors as the active ingredient in the O3 scrubber is depleted after long-term deployment due to cumulative exposure to O3. The net output from non-functional NO2 sensors at low ambient NO2 concentrations is thus a result of high-temperature effects and/or high-ozone effects, which can have variable influence over long deployment periods, resulting in the noisy behavior observed here. Considering the impact of temperature and ozone, sensor performance is expected to vary seasonally. Generally, sensors should perform better during winter with low temperature, low O3 concentration, and high NO2 concentration.

The influence of O3 concentration

In the previous section, the degradation pattern indicates that O3 scrubber depletion may be a major reason causing signals that significantly overestimate NO2 concentrations. To further investigate this cause, we analyzed the impact of ambient O3 concentrations on the NO2 sensors.

Figure 4 displays O3 influence changing over deployment duration. We tested the O3 influence by comparing two random forest (RF) models for inverting sensor signals. Each model was trained on 80% of the data and tested on the remaining 20%. For each collocation dataset, a baseline RF model (RFNO₂, O₃, RH, T) was trained with four input variables (NO2 and O3 raw signals, T, and RH). A second model was built that did not include O3 as an input (RFNO₂, RH, T). We then compared the difference in the R2 value of the two models for the testing data.

The color of data points in Figure 4 represents the R2 values of O3-excluded RF models. Data points in light yellow indicate that O3-excluded RF models can explain sensor raw signals well (R2 > 0.8), and data points in dark blue indicate the O3-excluded RF models cannot explain sensor signal variation. Most collocations that happened within 200 days after sensors unpacking demonstrate a light-yellow color in the lower-left corner of Figure 4, showing that the O3-excluded RF models can explain sensor performance well. These data points also have small y-values, indicating that the influence of O3 is negligible; the scrubber is still quite effective.

An upward trend can be observed when the collocation happened 200-400 days after sensor unpacking, and the color of datapoints gradually changes from light yellow to dark blue. This indicates that O3-excluded RF models cannot properly interpret signal variations, and adding O3 concentrations as input can significantly improve the baseline RF model. Such a trend shows that due to the depletion of O3 scrubbers, signals were biased by ambient O3 that was not removed effectively by the scrubber. Thus, when sensor aging started, including the O3 concentration in calibration models can partially address the issue and improve sensor performance.

When collocations happened more than 600 days after sensor unpacking, no noticeable trend can be observed. In this range, data points are scattered with dark blue colors, showing that ambient NO2 concentrations and environmental conditions can only partially explain signal variations, and the O3 influence is variable. Still, for some collocation datasets, the baseline RF models are significantly improved by including O3 concentrations, shown as blue points with large y-values. However, for other collocation datasets, including O3 concentrations cannot significantly improve baseline models, shown as blue points with small y-values. Low R2 values for both baseline and O3-excluded RF models could be due to sensor failure, which may happen 600 days after the sensors were unpacked. These failed sensors demonstrated R2 values around 0.2-0.3 for the baseline RF model. The manufacturer’s technical data sheet indicates that the operational lifetime of the NO2 sensor – defined as “months until 50% original signal” – is over 24 months (~800 days), but this means that the sensor gradually becomes less sensitive to both NO2 and O3 concentrations, which cannot be addressed by calibration models.

Chart, scatter chart

Description automatically generated

Figure 4. The impact of O3 concentration on NO2 sensors changes over time. R2NO₂, O₃, RH, T and R2NO₂, RH, T are the R2 values of the baseline RF models and the O3-excluded RF models. The difference between these two R2 values quantifies O3 influence, showing how ambient O3 concentration influences sensors’ response. Sensors usually started with high R2NO₂, RH, T values, shown as light-yellow points in the lower-left corner. An increasing trend can be seen within 400 days after unpacking, indicating that the O3 concentration gradually influences sensor response. After 400 days after sensor unpacking, no significant trend can be observed, showing that both NO2 and O3 concentrations can only partially explain sensor response.

With sensor aging, their detection limit increased gradually over deployment time, indicating worsening data quality (Figure S2). After 400 days, most sensors had a detection limit higher than the annual average NO2 concentration in Pittsburgh. Although sensor aging gradually increases the detection threshold, regular collocation and advanced algorithms can ease the aging issue.

Discussion and Recommendations

Sensor aging is an irreversible process, and identifying non-functional sensors in a timely manner is essential to ensure reliable data quality in long-term field deployments. Collocating sensors with reference instruments is the best way to check these sensors. We recommend estimating the lifetime of O3 scrubbers based on local O3 concentration and routinely collocating sensors with reference instruments after one year of deployment. If it is not possible to collocate sensors with reference instruments (usually defined as placing the sensor within ~10 m of the reference), it may be reasonable to check sensor response against a nearby reference monitor to obtain a rough evaluation of sensor performance, especially during low-NO2, high-O3 periods when concentrations across an urban network may be similar and comparable to a central monitoring station. Over long deployments (>2 years), the sensor likely becomes insensitive to the target species NO2; similar behavior may be seen in the electrochemical ozone sensor, but that is beyond the scope of this manuscript.

Strategically deploying sensors from the same batch may be an alternative option. For sensors from the same batch that are unpacked simultaneously, one or two sensors could be left at the reference site as an indicator, while other sensors can be deployed at different locations. By continuously examining the performance of the reference-site indicator sensor(s), one could monitor the aging of the batch and catch non-functional sensors quickly. However, this method can be biased as indicator sensor(s) and other sensors are exposed to different sources. Reference sites typically monitor regional data, and distributed sensors are often deployed near sources. Thus, compared to indicator sensor(s), other sensors may be exposed to a higher concentration that accelerates their aging.

Complicated advanced non-linear algorithms, for example, machine learning algorithms, may make it challenging to identify non-functional sensors. These advanced algorithms usually include temperature, relative humidity, and responses from other pollutant sensors as input variables. These other variables sometimes are naturally correlated with NO2 concentrations. Thus, these models make it challenging to determine whether the model-derived NO2 concentration is mainly determined by sensor raw signals or the other sensor inputs, which leads to the situation that the sensor still seems to report reasonable readings even after the sensor is non-functional. Directly correlating raw signals with reference NO2 concentrations would be a simple and practical solution to examine sensor status, whether during a collocation or alternative network-based approaches as described above.

Finally, as a conservative measure for similar high-ozone regions (e.g., Southern California and other urban areas, especially in the Global North), a network could simply replace NO2 sensors every six months, as performance degradation usually starts 200 days after sensor unpacking. Annual sensor replacement may suffice for cities (especially in the Global South) that experience lower ozone, e.g., Kigali, Rwanda,13 but this needs to be tested.

Supporting Information Available: The following files are available free of charge.

Supplemental Information\_text.docx: Reference sites information and sensors’ detection limit analysis.

Acknowledgments

This was developed as part of the Center for Air, Climate and Energy Solution (CACES). Funding was provided by the United States Environmental Protection Agency (assistance agreement nos. RD83587301 and RD83628601). It has not been formally reviewed by EPA. The views expressed in this document are solely those of authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. This work was also supported by the Heinz Endowment Fund (Grants E2375 and E3145). R.S. benefited from State assistance managed by the National Research Agency under the “Programme d’Investissements d’Avenir” under the reference “ANR-18-MPGA-0011” (“Make our planet great again” initiative).

**References**

1. (a) Chen, T.-M.; Kuschner, W. G.; Gokhale, J.; Shofer, S., Outdoor air pollution: nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects. *Am. J. Med. Sci.* **2007,** *333* (4), 249-256; (b) Samoli, E.; Aga, E.; Touloumi, G.; Nisiotis, K.; Forsberg, B.; Lefranc, A.; Pekkanen, J.; Wojtyniak, B.; Schindler, C.; Niciu, E., Short-term effects of nitrogen dioxide on mortality: an analysis within the APHEA project. *Eur. Respir. J* **2006,** *27* (6), 1129-1138.

2. (a) Andersen, Z. J.; Kristiansen, L. C.; Andersen, K. K.; Olsen, T. S.; Hvidberg, M.; Jensen, S. S.; Ketzel, M.; Loft, S.; Sørensen, M.; Tjønneland, A., Stroke and long-term exposure to outdoor air pollution from nitrogen dioxide: a cohort study. *Stroke* **2012,** *43* (2), 320-325; (b) Grazuleviciene, R.; Maroziene, L.; Dulskiene, V.; Malinauskiene, V.; Azaraviciene, A.; Laurinaviciene, D.; Jankauskiene, K., Exposure to urban nitrogen dioxide pollution and the risk of myocardial infarction. *Scand. J. Work Environ. Health* **2004**, 293-298; (c) Pershagen, G.; Rylander, E.; Norberg, S.; Eriksson, M.; Nordvall, S. L., Air pollution involving nitrogen dioxide exposure and wheezing bronchitis in children. *Int. J. Epidemiol* **1995,** *24* (6), 1147-1153; (d) Shima, M.; Nitta, Y.; Ando, M.; Adachi, M., Effects of air pollution on the prevalence and incidence of asthma in children. *Arch. Environ. Health* **2002,** *57* (6), 529-535; (e) Achakulwisut, P.; Brauer, M.; Hystad, P.; Anenberg, S. C., Global, national, and urban burdens of pediatric asthma incidence attributable to ambient NO2 pollution: estimates from global datasets. *Lancet Planet. Health***2019,** *3* (4), e166-e178; (f) Roberts, S.; Arseneault, L.; Barratt, B.; Beevers, S.; Danese, A.; Odgers, C. L.; Moffitt, T. E.; Reuben, A.; Kelly, F. J.; Fisher, H. L., Exploration of NO2 and PM2. 5 air pollution and mental health problems using high-resolution data in London-based children from a UK longitudinal cohort study. *Psychiatry Res.* **2019,** *272*, 8-17.

3. (a) Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R. P.; Lv, Q.; Hannigan, M., The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmos. Meas. Tech. Discuss.* **2014,** *7* (10), 3325-3336; (b) Zimmerman, N.; Presto, A. A.; Kumar, S. P.; Gu, J.; Hauryliuk, A.; Robinson, E. S.; Robinson, A. L.; Subramanian, R., A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech. Discuss.* **2018,** *11* (1), 291-313; (c) Tanzer-Gruener, R.; Li, J.; Robinson, A.; Presto, A., Impacts of Modifiable Factors on Ambient Air Pollution: A Case study of COVID-19 Shutdowns. *Environ. Sci. Technol. Lett.*, **2020,** *7*(8), 554-559; (d) Zimmerman, N.; Li, H. Z.; Ellis, A.; Hauryliuk, A.; Robinson, E. S.; Gu, P.; Shah, R. U.; Ye, Q.; Snell, L.; Subramanian, R., Improving correlations between land use and air pollutant concentrations using wavelet analysis: Insights from a low-cost sensor network. *Aerosol Air Qual. Res.* **2020,** *20* (2), 314-328; (e) Mijling, B.; Jiang, Q.; De Jonge, D.; Bocconi, S., Field calibration of electrochemical NO 2 sensors in a citizen science context. *Atmos. Meas. Tech. Discuss.* **2018,** *11* (3), 1297-1312; (f) Jiang, Q.; Kresin, F.; Bregt, A. K.; Kooistra, L.; Pareschi, E.; Van Putten, E.; Volten, H.; Wesseling, J., Citizen sensing for improved urban environmental monitoring. *J. Sens.* **2016.** DOI: doi.org/10.1155/2016/5656245; (g) Baron, R.; Saffell, J., Amperometric gas sensors as a low cost emerging technology platform for air quality monitoring applications: A review. *ACS Sens.* **2017,** *2* (11), 1553-1566; (h) Malings, C.; Tanzer, R.; Hauryliuk, A.; Kumar, S. P.; Zimmerman, N.; Kara, L. B.; Presto, A. A.; Subramanian, R., Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmos. Meas. Tech. Discuss.***2019,** *12* (2), 903-920; (i) Cross, E. S.; Williams, L. R.; Lewis, D. K.; Magoon, G. R.; Onasch, T. B.; Kaminsky, M. L.; Worsnop, D. R.; Jayne, J. T., Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. *Atmos. Meas. Tech. Discuss.***2017,** *10* (9), 3575-3588.

4. (a) Mueller, M.; Meyer, J.; Hueglin, C., 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.* **2017,** *10* (10), 3783-3799; (b) Jerrett, M.; Donaire-Gonzalez, D.; Popoola, O.; Jones, R.; Cohen, R. C.; Almanza, E.; De Nazelle, A.; Mead, I.; Carrasco-Turigas, G.; Cole-Hunter, T., Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. *Environ. Res.* **2017,** *158*, 286-294.

5. Mead, M.; Popoola, O.; Stewart, G.; Landshoff, P.; Calleja, M.; Hayes, M.; Baldovi, J.; McLeod, M.; Hodgson, T.; Dicks, J., The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmos. Environ.* **2013,** *70*, 186-203.

6. Stetter, J. R.; Li, J., Amperometric gas sensors a review. *Chem. Rev.* **2008,** *108* (2), 352-366.

7. (a) Smith, K. R.; Edwards, P. M.; Ivatt, P. D.; Lee, J. D.; Squires, F.; Dai, C.; Peltier, R. E.; Evans, M. J.; Sun, Y.; Lewis, A. C., An improved low-power measurement of ambient NO 2 and O 3 combining electrochemical sensor clusters and machine learning. *Atmos. Meas. Tech. Discuss.* **2019**, 1325-1336; (b) Esposito, E.; De Vito, S.; Salvato, M.; Bright, V.; Jones, R.; Popoola, O., Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sens. Actuators, B* **2016,** *231*, 701-713.

8. (a) Borrego, C.; Costa, A.; Ginja, J.; Amorim, M.; Coutinho, M.; Karatzas, K.; Sioumis, T.; Katsifarakis, N.; Konstantinidis, K.; De Vito, S., Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmos. Meas. Tech. Discuss* **2016,** *147*, 246-263; (b) Penza, M.; Suriano, D.; Villani, M. G.; Spinelle, L.; Gerboles, M. In *Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks*, SENSORS, 2014 IEEE, IEEE: 2014; pp 2012-2017.

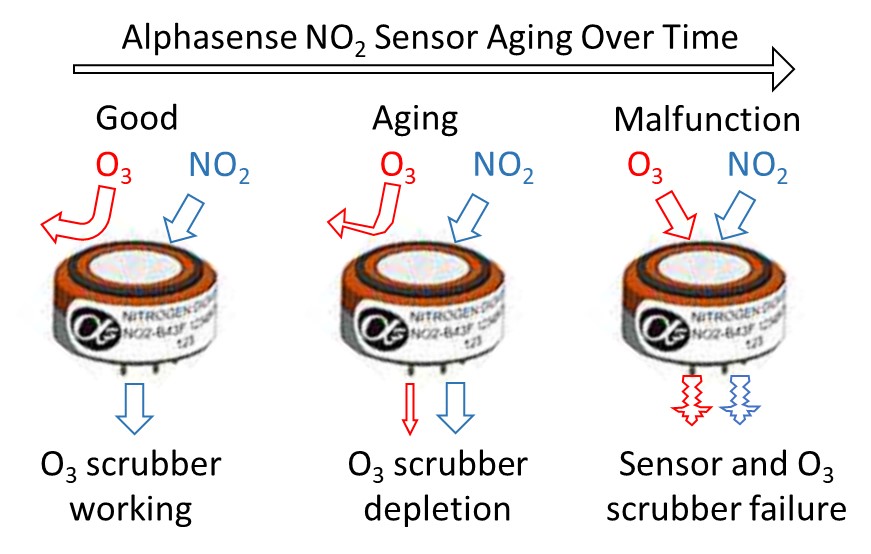
9. Sun, L.; Westerdahl, D.; Ning, Z., Development and evaluation of a novel and cost-effective approach for low-cost NO2 sensor drift correction. *Sensors* **2017,** *17* (8), 1916. DOI: 10.3390/s17081916

10. Hossain, M.; Saffell, J.; Baron, R., Differentiating NO2 and O3 at low cost air quality amperometric gas sensors. *ACS Sens.* **2016,** *1* (11), 1291-1294.

11. (a) Subramanian, R.; Ellis, A.; Torres-Delgado, E.; Tanzer, R.; Malings, C.; Rivera, F.; Morales, M.; Baumgardner, D.; Presto, A.; Mayol-Bracero, O. L., Air quality in puerto rico in the aftermath of hurricane maria: a case study on the use of lower cost air quality monitors. *ACS Earth Space Chem.* **2018,** *2* (11), 1179-1186; (b) Tanzer, R.; Malings, C.; Hauryliuk, A.; Subramanian, R.; Presto, A. A., Demonstration of a Low-Cost Multi-Pollutant Network to Quantify Intra-Urban Spatial Variations in Air Pollutant Source Impacts and to Evaluate Environmental Justice. *Int. J. Environ. Res. Public Health* **2019,** *16* (14), 2523. DOI: 10.3390/ijerph16142523; (c) Eilenberg, S. R.; Subramanian, R.; Malings, C.; Hauryliuk, A.; Presto, A. A.; Robinson, A. L., Using a network of lower-cost monitors to identify the influence of modifiable factors driving spatial patterns in fine particulate matter concentrations in an urban environment. *J. Exposure Sci. Environ. Epidemiol.* **2020**, 1-13.

12. (a) Abdi, H., The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA* **2007**, 508-510; (b) Kendall, M. G., A new measure of rank correlation. *Biometrika* **1938,** *30* (1/2), 81-93.

13. Subramanian, R.; Kagabo, A. S.; Baharane, V.; Guhirwa, S.; Sindayigaya, C.; Malings, C.; Williams, N. J.; Kalisa, E.; Li, H.; Adams, P., Air pollution in Kigali, Rwanda: spatial and temporal variability, source contributions, and the impact of car-free Sundays. *Clean Air J.* **2020,** *30* (2), 1-15.



**For TOC Only**