



Long-term calibration models to estimate ozone concentrations with a metal oxide sensor[☆]

Tofigh Sayahi ^{a,*}, Alicia Garff ^b, Timothy Quah ^c, Katrina Lê ^a, Thomas Becnel ^d, Kody M. Powell ^a, Pierre-Emmanuel Gaillardon ^d, Anthony E. Butterfield ^a, Kerry E. Kelly ^a

^a University of Utah, Department of Chemical Engineering, 3290 MEB, 50 S. Central Campus Dr., Salt Lake City, UT, United States

^b University of Utah, Department of Physics and Astronomy, 201 James Fletcher Building, 115 S. 1400 E, Salt Lake City, UT, United States

^c University of California, Santa Barbara, Department of Chemical Engineering, 3357 Engrg II, Santa Barbara, CA, United States

^d University of Utah, Department of Electrical and Computer Engineering, Laboratory for NanoIntegrated Systems, 50 S. Central Campus Dr., Salt Lake City, UT, United States

ARTICLE INFO

Article history:

Received 22 April 2020

Received in revised form

2 August 2020

Accepted 3 August 2020

Available online 12 August 2020

Keywords:

Ozone

Low-cost sensors

Metal oxide

Long-term calibration

ABSTRACT

Ozone (O₃) is a potent oxidant associated with adverse health effects. Low-cost O₃ sensors, such as metal oxide (MO) sensors, can complement regulatory O₃ measurements and enhance the spatiotemporal resolution of measurements. However, the quality of MO sensor data remains a challenge. The University of Utah has a network of low-cost air quality sensors (called AirU) that primarily measures PM_{2.5} concentrations around the Salt Lake City valley (Utah, U.S.). The AirU package also contains a low-cost MO sensor (\$8) that measures oxidizing/reducing species. These MO sensors exhibited excellent laboratory response to O₃ although they exhibited some intra-sensor variability. Field performance was evaluated by placing eight AirUs at two Division of Air Quality (DAQ) monitoring stations with O₃ federal equivalence methods for one year to develop long-term multiple linear regression (MLR) and artificial neural network (ANN) calibration models to predict O₃ concentrations. Six sensors served as train/test sets. The remaining two sensors served as a holdout set to evaluate the applicability of the new calibration models in predicting O₃ concentrations for other sensors of the same type. A rigorous variable selection method was also performed by least absolute shrinkage and selection operator (LASSO), MLR and ANN models. The variable selection indicated that the AirU's MO oxidizing species and temperature measurements and DAQ's solar radiation measurements were the most important variables. The MLR calibration model exhibited moderate performance ($R^2 = 0.491$), and the ANN exhibited good performance ($R^2 = 0.767$) for the holdout set. We also evaluated the performance of the MLR and ANN models in predicting O₃ for five months after the calibration period and the results showed moderate correlations (R^2 s of 0.427 and 0.567, respectively). These low-cost MO sensors combined with a long-term ANN calibration model can complement reference measurements to understand geospatial and temporal differences in O₃ levels.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Ground level ozone (O₃) is produced by chemical reactions of gaseous precursors in the presence of sunlight. The highly reactive nature of O₃ makes this gas a potent oxidant, detrimental to human health, vegetation, and ecosystem productivity. Epidemiological studies show that O₃ is a significant risk factor for increased mortality, exacerbations of respiratory diseases and increased hospital

admissions (Anenberg et al., 2010; Cooper et al., 2019; Sousa and Martins, 2013). Government organizations typically measure O₃ concentrations from sparsely distributed air quality monitoring stations equipped with highly accurate regulatory instruments. However, these instruments are expensive and cumbersome to maintain, restricting the number of instruments that can be deployed. Consequently, these instruments are unable to accurately capture the spatiotemporal variability of O₃ concentrations in urban areas (Kouvarakis, 2002; Park and Kwan, 2017; Tzortziou et al., 2015). One strategy to enhance the geospatial resolution of O₃ measurements is by adding low-cost sensors to complement the existing conventional monitoring networks. A network of low-cost

[☆] This paper has been recommended for acceptance by Charles Wong.

* Corresponding author.

E-mail address: tofighsayahi@chemeng.utah.edu (T. Sayahi).

electrochemical or gas-sensitive semiconductor sensors (two common inexpensive strategies for estimating/inferring ozone) would provide higher temporal and spatial resolution (Schneider et al., 2017).

One of the major challenges with using low-cost O₃ sensor networks is the accuracy of the sensor data in deployed conditions. Most of the low-cost sensor manufacturers do not calibrate the sensors or if they do, the calibration takes place in a laboratory chamber which may not reflect the behavior of the sensors under real-world conditions (Castell et al., 2017; Ferrer-cid et al., 2019). Therefore, the sensor technical information that the manufacturers provide is often not sufficient, and field calibration is needed to ensure the suitability of the sensors for the intended application. Field calibration of the low-cost O₃ sensors has some limitations. The sensors may exhibit cross-sensitivity with other ambient pollutants, such as NO₂ (Mead et al., 2013; Peterson et al., 2017). Moreover, these low-cost sensors may also be affected by meteorological factors, including relative humidity (RH) and temperature (Hitchman et al., 1997; Masson et al., 2015; Peterson et al., 2017). Another limitation associated with these low-cost sensors is that they may drift over time (Masson et al., 2015).

Since the deconvolution of the effects of meteorological factors, cross-sensitivity, and stability on the performance of the low-cost sensors is complex, there has been an increased interest in using sophisticated calibration methods, such as computational modeling. Many studies (Hagan et al., 2018; Malings et al., 2019; Spinelle et al., 2015; Zimmerman et al., 2018) collocated low-cost sensors with a reference monitor for a few weeks and used these advanced methods to establish calibration models for the low-cost sensors. The choice of calibration model depends on the behavior of the low-cost sensor response in comparison to the collocated reference monitor. Multiple linear regression (MLR) models are used when the sensor shows a linear response in comparison to a collocated reference monitor (Hagan et al., 2018). Barcelo-ordinas et al. (2019) calibrated 136 MO sensors by collocating them with reference ozone monitors in Spain and Italy over the time period of three to four weeks using the MLR method. The results indicated good calibration estimates (root mean square error, RMSE, of 3.97–12.2 ppb); however, the model produced less accurate predictions over the long-term (three months, RMSE up to 20.4–22.9 ppb). When the low-cost O₃ sensor shows a nonlinear behavior, more complex machine learning models such as artificial neural network (ANN, Spinelle et al., 2015), K-nearest neighbors (Hagan et al., 2018), Gaussian process regression (GPR; Malings et al., 2019), random forest (RF, Zimmerman et al., 2018), and support vector regression (SVR; Esposito et al., 2016) have been implemented to calibrate the sensors. Esposito et al. (2016) performed air pollution field calibration (for NO, NO₂, and O₃) using ANN and SVR and compared these models to MLR and GPR using four weeks of data. Their results demonstrated that in the case of nonlinearity in the sensors responses, SVR and ANN performed better than MLR and GPR at the expense of higher computational resources.

One of the main challenges with the current sensor calibration models for O₃ is that the time period of the training phases is limited, i.e., few weeks. Therefore, computational models developed in previous studies show high uncertainties in predicting long-term O₃ concentrations, likely in part because the meteorological conditions change over time (Malings et al., 2019; Spinelle et al., 2015; Zimmerman et al., 2018). Moreover, it is cumbersome to co-locate numerous low-cost O₃ sensors with a reference monitor in the field and establish a calibration model for each sensor; therefore, another key challenge is whether a model that is developed to calibrate an O₃ sensor can be used to calibrate other sensors of the same type. This study addresses these challenges by

developing two generic calibration models (MLR and ANN) that consider intra-sensor variability utilizing six inexpensive MO sensors (\$8) over an extended period of time (one year) as well as assessing the capability of the models to apply to other sensors of the same type. Developing a long-term calibration model applicable to multiple sensors of the same type facilitates the calibration of large number of low-cost sensors before deploying them in a network of sensors.

2. Materials and methods

The University of Utah has a working network of more than 100 low-cost air quality sensors Salt Lake City valley (Utah, the United States of America) using a sensor package named AirU (Becnel et al., 2019b; Kelly et al., 2017; Sayahi et al., 2019a, 2019b). This network focuses primarily on gathering and communicating particulate matter concentrations (AQ&U, 2019). The AirU package includes an inexpensive MO sensor (\$8), specifically designed to detect NO₂ and CO, although the NO₂ and CO sensors are capable of measuring oxidizing (OX_{AirU}) and reducing (RD_{AirU}) species, respectively. In this study, we first evaluated the effectiveness of this MO sensor in measuring O₃ (because O₃ is also a strong oxidant) in laboratory settings using simple models such as linear and Langmuir models. Then, we collocated eight AirU packages with reference monitors in the field at two state monitoring stations over a one-year period. Before applying any sophisticated method, we utilized linear and Langmuir models to calibrate the eight MO sensors. Then, we used the measurements from six AirU packages as well as the reference monitors during the one-year calibration period to develop long-term MLR and ANN calibration models for the AirU packages to predict O₃ concentrations. The data gathered from the other two sensors were used to assess the applicability of the developed calibration models in estimating ozone concentrations for other sensors of the same type. We also assessed the performance of the two models for 5 months after the calibration period. Table 1 summarizes the nomenclature and variables used in this study.

2.1. AirU package

The AirU package integrates several meteorological and air quality sensors into a single package. Airborne particulate matter is measured with a Plantower PMS3003 optical light-scattering particle counter. The PMS3003 reports PM₁, PM_{2.5}, and PM₁₀ (particulate matter less than 1, 2.5, and 10 μm and denoted by PM_{1,AirU}, PM_{2.5,AirU}, and PM_{10,AirU}, respectively) concentrations in μg/m³. The values are sent from the PMS3003 to the Espressif ESP32 microcontroller (MCU) at approximately 1-s intervals via the Universal Asynchronous Receiver-Transmitter (UART) communication protocol. The AirU incorporates a Texas Instruments HDC1080 digital temperature (T_{AirU}) and RH (RH_{AirU}) sensor, and the MCU periodically (every minute) polls temperature and RH data using the Inter-Integrated Circuit (I2C) communication protocol. The AirU also incorporates a Quectel L72 GPS module for global positioning and atomic time synchronization via the GPS Real-time Clock.

The AirU also collects oxidizing and reducing gas concentrations, and reports them as NO₂ and CO readings, using an SGX MiCS 4514 MO sensor, which is an n-type SnO₂ semiconductor sensor. The operating principle behind the MiCS 4514 relies on two sensing elements doped with a tin oxide variant that attracts and reacts to either oxidizing or reducing species. Reducing gasses that collect on the sensing element will reduce the resistance of the element. Conversely, oxidizing gasses that collect on the sensing element will increase the resistance of the element. The two sensing elements are connected to two independent voltage divider circuits. The outputs of the voltage dividers are buffered by preamplifiers

Table 1
The nomenclature and variables.

Abbreviation	Definition
ADC	Analog-to-digital converter
AirU	University of Utah sensor package
ANN	Artificial neural network
AQ&U	University of Utah network of low-cost air quality sensors
B	Bias
c_1, c_2 , and c_3	Langmuir model constants
DAQ	Utah Division of Air Quality
EPA	Environmental Protection Agency
GPR	Gaussian process regression
I2C	Inter-integrated circuit
LASSO	Least absolute shrinkage and selection operator
LOD	Limit of detection
MLR	Multiple linear regression
MO	Metal oxide
MSE	Mean squared error
N	Number of data points
N-FET	N-channel metal-oxide field-effect transistor
O_3	Ozone
$O_{3,DAQ}$	DAQ ozone
$O_{3,DAQ,i}$	True ozone value
$O_{3,DAQ,t}$	True ozone value at time t
$O_{3,pred,i}$	Predicted ozone output value
$O_{3,pred,t}$	Predicted ozone output value at time t
OX_{AirU}	AirU oxidizing species
P	Precision
$PM_{1,AirU}$	Particulate matter less than 1 μm
$PM_{10,AirU}$	Particulate matter less than 10 μm
$PM_{2.5,AirU}$	Particulate matter less than 2.5 μm
RD_{AirU}	AirU reducing species
RF	Random forest
RH_{AirU}	AirU relative humidity
RH_{DAQ}	DAQ relative humidity
RMSE	Root mean square error
SD	Standard deviation
SR_{DAQ}	DAQ solar radiation
SVR	Support vector regression
T_{AirU}	AirU temperature
T_{DAQ}	DAQ temperature
UART	Universal asynchronous receiver-transmitter

before being fed into two channels of a Texas Instruments ADS1015 12-bit delta-sigma analog-to-digital converter (ADC) integrated circuit. The ADS1015 provides 1 milli-Volt noise-free resolution with a 5 V analog reference and a low-pass filter with a cutoff frequency of 10 kHz. 12-bit values (0–4096) are read from the two ADS1015 channels via I2C by the ESP32 host MCU, where they are either stored or sent to an online database. The reducing circuit uses an external 82 Ohm resistor in the voltage divider, and the oxidizing circuit uses an external 82 Ohm resistor in series with a 51 Ohm resistor, for a total of 133 Ohms of resistance. As per the manufacturer's recommendations, an N-Channel Metal-Oxide Field-Effect Transistor (N-FET) is added as a heater in parallel with the oxidizing circuit 51 Ohm resistor. When the N-FET is activated, the 51 Ohm resistor is shorted, and the equivalent external resistance is 82 Ohms. By decreasing the resistance, the current is increased, and thus the sensor temperature is also increased. This heater is used for cleaning purposes by increasing the reaction rate of the accumulated gas (Peterson et al., 2017). All data collected by the AirU can either be stored on an on-board micro-SD card, or published to an online database using the ESP32 MCU WIFI coprocessor. Details of the AirU package can be found in (Becnel et al., 2019a, 2019b).

2.2. Laboratory evaluation

The objective of this component of the study is to evaluate the ability of the MO sensors to measure ozone concentration under

controlled laboratory conditions. Fig. S-1 illustrates the laboratory setup. A 2B Technologies™ ozone calibration source (model 306) generated O_3 and a 2B Technologies™ ozone monitor (model 106-L) measured O_3 as the reference monitor. We evaluated three MO sensors, each integrated into an AirU, for their responses to ozone and for the effects of environmental factors including RH and temperature. We achieved a low temperature (average of 13.8 °C) by testing inside a Danby freezer (model DCFM050C1) and a high temperature (average of 40.8 °C) using a seedling heating mat (NAMOTEK 120 V) inside the freezer (while the freezer was turned off). RH (28.2%–57.2%) was adjusted using an ultrasonic atomizer.

Before the start of each test, the MO sensors on the AirUs were placed in the chamber and allowed to warm up for 30 min by activating the on-board heater. After the warm up period, a fan provided a well-mixed environment (Fig. S-1). During each test, the ozone generator was set at 700 ppb, which resulted in a concentration of approximately 100 ppb within the chamber, and provided ozone to the chamber for three 30-min periods. Each 30-min period of ozone generation was followed by a 20-min period of “zero air” generation where the inlet airflow of O_3 was 0 ppb as measured by the 2B reference monitor. All tests lasted a total of 2 h and 30 min, excluding the warm up time. Linear and Langmuir isotherm models were used to fit the laboratory OX_{AirU} measurements to O_3 concentrations. The Langmuir fit (Eq. 1) was chosen because it describes the physical process of the gas adsorption-desorption interactions with the MO sensors (Gomri et al., 2005):

$$O_{3,DAQ} = \frac{c_1(OX_{AirU} + c_3)}{c_2 + OX_{AirU}} \quad (1)$$

where c_1 , c_2 , and c_3 are the Langmuir model constants and $O_{3,DAQ}$ is the O_3 reference monitor readings.

2.3. Field evaluation

In order to assess the performance of the MO sensors in measuring O_3 concentrations under real-world conditions, four AirU sensors at each of two locations (for a total of eight sensors) were co-located with an ozone federal equivalent method (FEM, Teledyne API T400 Ozone Analyzer) at two Utah Division of Air Quality (DAQ) monitoring stations for a time period of one year (October 1st 2018 to September 30th 2019). The two DAQ stations (Fig. S-2), Hawthorne (AQS: 49-035-3006, longitude: –111.8722, latitude: 40.7364, elevation: 1312 m) and Rose Park (AQS: 49-035-3010, longitude: –111.9310, latitude: 40.7842, elevation: 1292 m), are located 8.3 km apart in urban residential areas of Salt Lake City, Utah, USA. One-minute readings from each of the AirU's MO sensors were averaged on an hourly basis to compare to the O_3 FEM monitor. The on-board heater operated for 30 min twice a day (7:30 and 15:30 mountain standard time) to clean the MO sensor. Consequently, the 7:00 and 15:00 h were excluded from this study. Similar to the laboratory study, linear and Langmuir isotherm models were chosen to evaluate the ability of MO sensors to measure O_3 levels.

This study focused on measurements that are likely to be correlated with O_3 concentrations and/or the MO sensor's measurement of O_3 , including the AirU package measurements of time, $PM_{1,AirU}$, $PM_{2.5,AirU}$, $PM_{10,AirU}$, OX_{AirU} , RD_{AirU} , T_{AirU} , RH_{AirU} as well as the DAQ measurements, such as solar radiation (SR_{DAQ} ; Eppley Radiometer, model 8-48), temperature (T_{DAQ}) and RH (RH_{DAQ}). It should be noted that Rose Park station does not measure solar radiation and in developing this model we assumed that SR is the same throughout the study area. As a first step, we conducted a cross correlation using Spearman rank correlation in Python 3 to evaluate the pairwise linear relationships between the $O_{3,DAQ}$ and

each of the abovementioned measurements.

2.4. Modeling study

The objective of this part of the study is to develop a long-term calibration model to predict O_3 using the measurements available in the AirU package and DAQ stations (Section 2.3). This model can then be applied to the AirU sensors in the AQ&U network to provide a high-resolution temporospatial ozone map for the Salt Lake valley. Sensor 2 and sensor 6 were randomly selected as the holdout set to evaluate the applicability of the developed model in calibrating other sensors of the same type. The other six sensors were used to train and test the model; 70% of the data were randomly selected as the training set and the remaining 30% as the test set.

2.4.1. Independent input variable selection

In order to optimize the number of variables used in developing the calibration model, least absolute shrinkage and selection operator (LASSO) and MLR models were developed on the collected measurements using Sklearn standard package of Python 3. The details of these two models can be found in Goldberger (1964) and Tibshirani (1996). RMSE and R^2 metrics were used to compare the goodness of fits for the models. LASSO regression is a type of linear regression that reduces the multicollinearity among the independent variables by penalizing the absolute value of the magnitudes of the variable coefficients. It basically keeps the most important parameters in a model by shrinking all the coefficient of similar independent variables to zero. Before training the measurements in the LASSO model, the alpha variable (LASSO hyperparameter) was tuned to 0.0001 by applying a grid search with a 5-fold cross-validation strategy to the train/test sets. In the MLR method, the number of input variables used to predict the actual O_3 concentrations ($O_{3,DAQ}$) were plotted against the MLR model metrics to find the optimum number of independent variables. A 5-fold cross-validation method was implemented in the MLR model to maximize the number of measurements used for training and testing the model. To ensure that the variables selected with MLR and LASSO models were the optimum variables for the calibration model, the metrics of the developed machine learning model (discussed in Section 2.4.2) were also plotted against the number of independent variables similar to the MLR approach.

2.4.2. Model development

The objective of this part of the study is to develop an MLR and supervised ANN models to predict O_3 concentrations using the independent variables selected in Section 2.4.1. The MLR model was implemented with a 5-fold cross-validation method similar to the previous section. The supervised ANN model used in this study is a multilayer perceptron model, which is a class of feed-forward neural network, with a minimum of three layers (including one hidden layer). Details of the MLP method can be found in (Hsieh, 2009).

Computational resources from Google Collaboratory (Jupyter Notebook with GPU and TPU resources) and Kaggle (Jupyter Notebook with GPU resources) were used to develop this model. The ANN model was implemented in Python 3 using Keras as a frontend for TensorFlow (Abadi et al., 2016). The parameter space for the network's nodes and hidden layers varied from 6 to 100 nodes and from 1 to 4 hidden layers. The network's weights were changed to minimize the difference between the model's prediction of a value and the true value (loss), and after several iterations the trained network was returned. The ANN model's hyperparameters included the loss metric, optimizer, epoch, batch size, and drop rate. The selected loss metric was the mean squared error (MSE):

$$MSE = \frac{1}{N} \left(\sum_i^N (O_{3,pred,i} - O_{3,DAQ,i})^2 \right) \quad (2)$$

where N is the number of data points, and $O_{3,pred,i}$ and $O_{3,DAQ,i}$ are the predicted and true output values, respectively.

The Adam optimization method was chosen to minimize loss (MSE) by changing the weight matrix (Kingma and Ba, 2015) using a stochastic optimizer implemented in Keras. The parameters related to the Adam optimizer followed the default settings except for the learning rate, which was set to 0.05. The number of training cycles (epochs) was 100. Both the batch size and dropout rate were implemented to prevent overfitting of the training data. The batch size, which determines how many data points are used to train before updating the network, was set to 32. The dropout rate, which is the percent of nodes that are randomly selected to be removed, was set to be 20%.

2.4.3. Model performance criteria

This section evaluates the performance of the sensors/calibration model with performance criteria for low-cost air quality measurements developed by the Environmental Protection Agency (EPA, Williams et al., 2014). Specifically, we averaged the prediction and actual O_3 data every 8 h and calculated the bias (B) and precision (P) of the MLR and ANN calibration models for predicting the O_3 concentrations for the holdout sensors. It should be noted that we used the 8-h averaged measurements only for Section 3.4.3 (the rest of Section 3 is based on hourly averaged measurements), and we excluded the measurements associated with 7:00 and 15:00 h. We calculated the bias (B, (Eq. 3)) for each model as the mean of the ratio of the O_3 prediction at time t ($O_{3,pred,t}$) to the corresponding actual O_3 measurement at time t ($O_{3,DAQ,t}$):

$$B_t = \left(\frac{PMS_t}{Ref_t} - 1 \right) \times 100 \quad B = \text{mean} \left(\frac{O_{3,pred,t}}{O_{3,DAQ,t}} - 1 \right) \times 100 \quad (3)$$

We excluded the $O_{3,DAQ,t}$ measurements less than 1 ppb from the calculation of sensor bias. We also estimated the precision (P, (Eq. (4)) of each model as the ratio of standard deviation for all O_3 predictions for each model, SD ($O_{3,pred}$), over the mean of all O_3 actual values, mean ($O_{3,DAQ}$):

$$P = \left(\frac{SD(O_{3,pred})}{\text{mean}(O_{3,DAQ})} \right) \times 100 \quad (4)$$

2.4.4. Post-calibration assessment

With the aim to evaluate the performance of the MLR and ANN models after the calibration period, we continued collocating the 8 AirU packages with the FEM instruments at the two DAQ stations. In this study, we used the developed models to predict 5 months of O_3 measurements from October 1st, 2019 to February 29th, 2020.

3. Results and discussion

3.1. Laboratory evaluation

This portion of the study assessed the capability of the MO sensors in measuring O_3 concentrations under controlled settings as a screening metric. Fig. S-3 fits the OX_{AirU} measurements from the three AirU sensors with the O_3 reference monitor measurements using linear and Langmuir models. Each figure includes three distinct subtests. These subtests correspond to the three different maxima. The AirU MO sensors follow the O_3

concentrations trends from the collocated reference monitor through the three subtests (Fig. S-3-A). Although both the linear and Langmuir fits represent the behavior of AirU MO sensors in measuring O_3 concentrations under all test conditions, the Langmuir model provided a better fit (Tables S-1, linear: R^2 of 0.680–0.937, Langmuir: R^2 of 0.699–0.975). Other studies support the use of a Langmuir isotherm as a model for gas-adsorption by MO sensors (Gomri et al., 2006, 2005; Hua et al., 2018). The laboratory results also revealed intra-sensor variability (Tables S-1), which was also reported by other studies of MO sensor performance (Collier-Oxandale et al., 2020; Gomri et al., 2005; Malings et al., 2019; Masson et al., 2015; Vito et al., 2018). This variability can be partly attributed to the differences in the semiconductor characteristics of the MO sensors of the same type (Masson et al., 2015).

Two environmental factors (temperature and RH) affected the laboratory model fits (Tables S-1), with better fits (RMSE and R^2) at higher temperatures and lower RH. Temperature affects the sensors' conductivity and resistance. An oxidizing gas, such as O_3 , removes free electrons from the conduction band, decreasing the conductivity of the sensor and increasing its resistance (Becker et al., 1999). Temperature is directly related to the reaction rate of O_3 binding to our n-type semiconductor sensor. RH also has an impact on the sensors' measurements. When water molecules are present on the surface of the sensor, the gas molecules are less likely to reach the surface of the sensor (Sohn et al., 2008).

Based on these screening laboratory results, the MO sensors in the AirU are capable of measuring O_3 . These findings support the large body of research that has validated metal oxide sensors as good sensors for O_3 (Becker et al., 1999; Mueller et al., 2017; Peterson et al., 2017; Ripoll et al., 2019; Vito et al., 2018). The promising results of this study suggest that the existing network of sensors in the Salt Lake valley (AQ&U) may be able to measure O_3 in the field.

3.2. Ambient evaluation

The promising laboratory performance using simple linear and Langmuir isotherm models, described in Section 3.1, was not observed when the sensors operated in the ambient environment. Tables S-2 and Fig. S-4 show the comparison between the OX_{AirU} readings from eight AirU sensors collocated with a reference monitor at two state monitoring stations for a one-year time period. In this field evaluation, fitting the OX_{AirU} resistance from the MO AirU sensor with both linear and Langmuir models yielded poor predictions of O_3 concentration (R^2 ranges of 0.017–0.177 and 0.168–0.256, respectively). Similar to the laboratory tests, the field evaluation also demonstrated some intra-sensor variability.

The laboratory evaluation suggested that environmental factors could affect the resistance measured by the oxidizing species MO sensor. The results of the pairwise cross correlations between the O_3 reference measurements and other environmental measurements are summarized in Tables S-3. The positive sign indicates a positive correlation while a negative sign indicates a negative correlation. Higher absolute values show higher correlations between the variables. The results show that generally for the low-cost sensors, both RD_{AirU} and OX_{AirU} have positive relationships with the O_3 measurements (except AirU 8 for the reducing species). Since O_3 is an oxidizing agent we would expect a positive correlation with the oxidizing species and a negative correlation with the reducing species; however, for the RD_{AirU} , the low-cost MO sensors that we used in this study (except AirU 8) did not operate as anticipated.

The temperature (both DAQ and AirU), RH and SR have relatively high correlations with O_3 concentrations. The discussion about the

effect of temperature and relative humidity can be found in Section 3.1. It should be noted that T_{AirU} has a high linear correlation with T_{DAQ} ($T_{DAQ} = 1.05T_{AirU} + 9.59$, $R^2 = 0.738$) while RH_{AirU} has a moderate linear relationship with RH_{DAQ} ($RH_{DAQ} = 0.758RH_{AirU} + 27.6$, $R^2 = 0.475$). Regarding SR, researchers have shown that more sunlight (higher SR) and its inherent higher temperature lead to higher ground-level O_3 concentrations (Jackman et al., 2001; Solberg et al., 2008). Negative correlations were also found between O_3 and PM concentrations. Similar to the effect of RH, the presence of particles on the sensor may inhibit the ability of O_3 to reach the sensor surface.

3.3. Independent input variable selection for the calibration model

We employed LASSO, MLR, and ANN models to select input variables for the training set to develop a simple calibration model and lower the risk of overfitting. Recognizing the challenge of intra-sensor variability (Section 3.1), we evaluated the broader applicability of the calibration model by selecting six sensors for the training/testing sets and two sensors for the holdout set.

Applying LASSO regression was the first step in selecting the most important variables for our calibration models. Fig. 1 shows the magnitudes of the LASSO coefficients for all 11 potential input variables, and it reveals that OX_{AirU} package and SR_{DAQ} are the only non-zero coefficients. The selection of OX_{AirU} and SR_{DAQ} as the most important input variables in estimating the O_3 concentrations seems physically reasonable as discussed in Section 3.2. By choosing SR_{DAQ} as one of the main factors to estimate O_3 , we assume that solar radiation is the same everywhere in the study area since it is only measured in Hawthorne State Monitoring station. The LASSO model with the two selected input variables of OX_{AirU} and SR_{DAQ} were moderately correlated to the O_3 measurements with an R^2 of 0.644 and an RMSE of 10.5 ppb. It should be noted that even though LASSO regression does not consider time as an important variable in improving the model, we use time as the third variable in both our ANN and MLR models to include the O_3 seasonal variation in the model.

We applied the MLR model to determine the optimum number of input variables for the calibration model (Fig. 2). We started with all the potential input variables (11) and eliminated the variables one by one without replacement (from right to left side of each subfigure in Fig. 2) in the following order: RD_{AirU} , RH_{AirU} , RH_{AirU} , $PM_{10,AirU}$, $PM_{2.5,AirU}$, $PM_{1,AirU}$, T_{DAQ} , T_{AirU} , OX_{AirU} , and SR_{DAQ} . It should be noted that we determined the order of variable elimination by

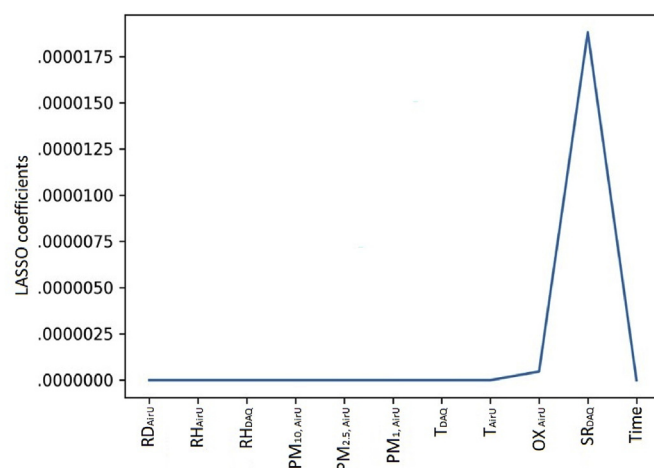


Fig. 1. The magnitudes of LASSO coefficients for eleven potential input variables.

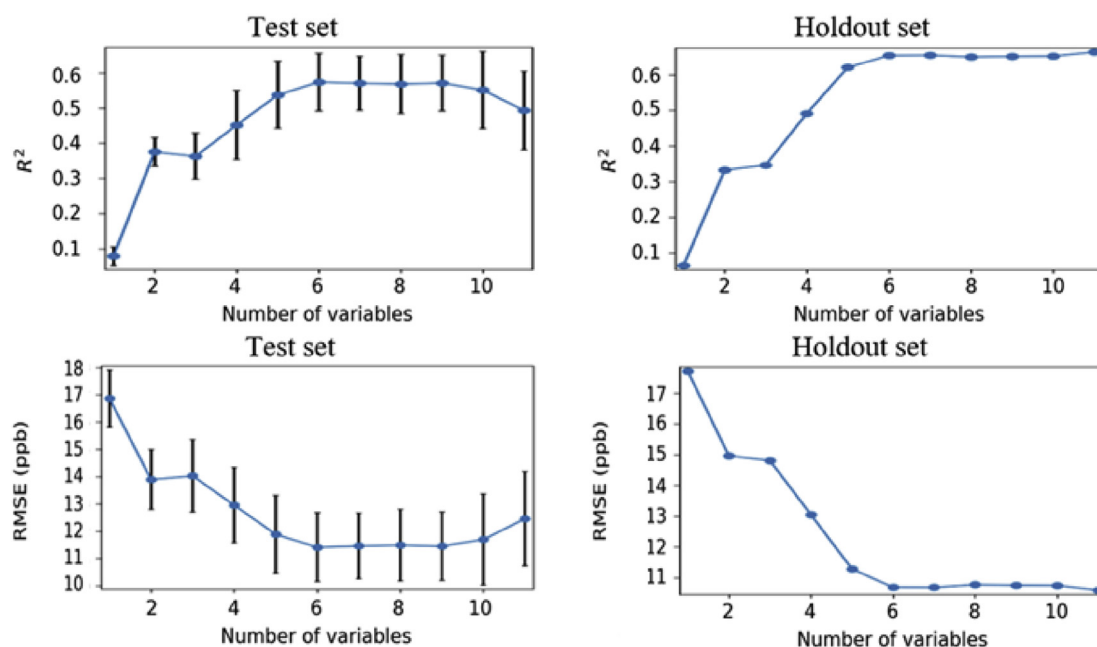


Fig. 2. Goodness of fit metrics (R^2 and RMSE) for the MLR model variable selection. The errorbars for the test set indicate the standard deviation of R^2 and RMSE associated with the 5-fold cross-validation. There are no errorbars for the holdout set because the model was applied to the entire dataset. It should be noted that the variables were eliminated in the following order: RD_{AirU} , RH_{AirU} , $PM_{10,AirU}$, $PM_{2.5,AirU}$, $PM_{1,AirU}$, T_{DAQ} , T_{AirU} , OX_{AirU} , and SR_{DAQ} . This elimination was applied without replacement from right to left side of each subfigure. On the right side of the subfigures, the model was applied for 11 variables then the first variable (RD_{AirU}) was eliminated and the model was applied for 10 variables. Then, the second variable (RH_{AirU}) was eliminated and the model was applied for 9 variables so on and so forth.

omitting each variable (with replacement) and implementing the MLR model (Fig. S-5). Better goodness of fit metrics (higher R^2 and lower RMSE) for an omitted variable means that the variable had negative effect on the model response. RD_{AirU} was eliminated first because for seven (out of eight) sensors, it was not operating as we expected (see Section 3.2) and had a large intra-sensor variability (Tables S-3). The temperature of the station was eliminated before the temperature of the AirU sensor because temperature varies in different areas that AQ&U network covers at Salt Lake valley and we want to develop a model based on the specific conditions of each sensor. Moreover, the temperature inside the sensor housing can be warmer than outside, particularly if it is located in direct sunlight.

In Fig. 2, the MLR goodness of fit metrics are plotted against the number of variables used to predict O_3 concentration. The results suggest that the optimum number of input variables is five, which means in addition to the time and the two variables that LASSO model suggested (OX_{AirU} and SR_{DAQ}), T_{DAQ} and T_{AirU} are also needed. However, to avoid overfitting, we chose to add only T_{AirU} as the fourth input variable since, as mentioned before, it is linearly correlated with T_{DAQ} .

To implement the ANN model, we first applied Adam optimization to all 11 variables to determine the optimum number of nodes and hidden layers (Fig. S-6). The results show that a shallow neural network with one hidden layer and 25 nodes would provide a simple model with high goodness of fit metrics. Fig. S-7 shows the predicted O_3 values against the actual O_3 concentration (parity plot) for all the variables using one hidden layer and 25 nodes (R^2 of 0.821 ± 0.00562 and RMSE of 0.487 ± 0.007 ppb on the training and R^2 of 0.820 ± 0.0120 and RMSE of 0.592 ± 0.213 ppb on the holdout set). Then, similar to the MLR method, we conducted the variable elimination process (without replacement) using the ANN model (Fig. S-8). The ANN results also suggest that adding T_{AirU} to the input variables that LASSO recommended is beneficial to obtain a better performance of the calibration model in estimating O_3 levels. Therefore, we use the following four variables to predict O_3

concentrations using ANN and MLR models: time, SR_{DAQ} , OX_{AirU} , and T_{AirU} . These four variables have demonstrated associations with O_3 concentrations (Becker et al., 1999; Hagler et al., 2018; Sohn et al., 2008; Solberg et al., 2008).

3.4. Evaluation of the calibration models

We developed both linear (MLR) and non-linear (ANN) calibration models to predict O_3 concentrations using one year of measurements from the four selected input variables described in Section 3.3.

3.4.1. MLR model

Fig. 3 shows the parity plots of the test and holdout sets for the MLR model. The results demonstrate that this model has moderate correlations (R^2 of 0.452 ± 0.0978 for test set and 0.491 for holdout set) and high accuracies (RMSE of 13.0 ± 1.38 ppb for test set and 0.0130 ppb for holdout set) in predicting O_3 concentrations. Low to moderate correlations of a MLR calibration model with O_3 concentrations estimated using MO sensors are also reported by other studies (Barcelo-ordinas et al., 2019; Ripoll et al., 2019; Vito et al., 2018; Zimmerman et al., 2018). For example, Ripoll et al. (2019) calibrated 132 Captor MO sensors in using MLR model by collocating them with O_3 reference monitors at air quality monitoring stations in Italy and Spain for two weeks, and the results showed that 88 sensors (67% of total) had R^2 between 0.1 and 0.6. This behavior could be because MO sensors generally underestimate high O_3 levels and have non-linear responses (Lung et al., 2018; Moltchanov et al., 2015; Ripoll et al., 2019; Spinelle et al., 2015).

Fig. S-9 shows the O_3 concentration and the MLR prediction for the holdout of sensors over the first week of July 2019 when there was a large variation in O_3 concentrations. The results indicate that the MLR model is responsive to the large changes of O_3 concentrations and follows its diurnal patterns: increasing in the morning, peaking around the midday and decreasing in the afternoon.

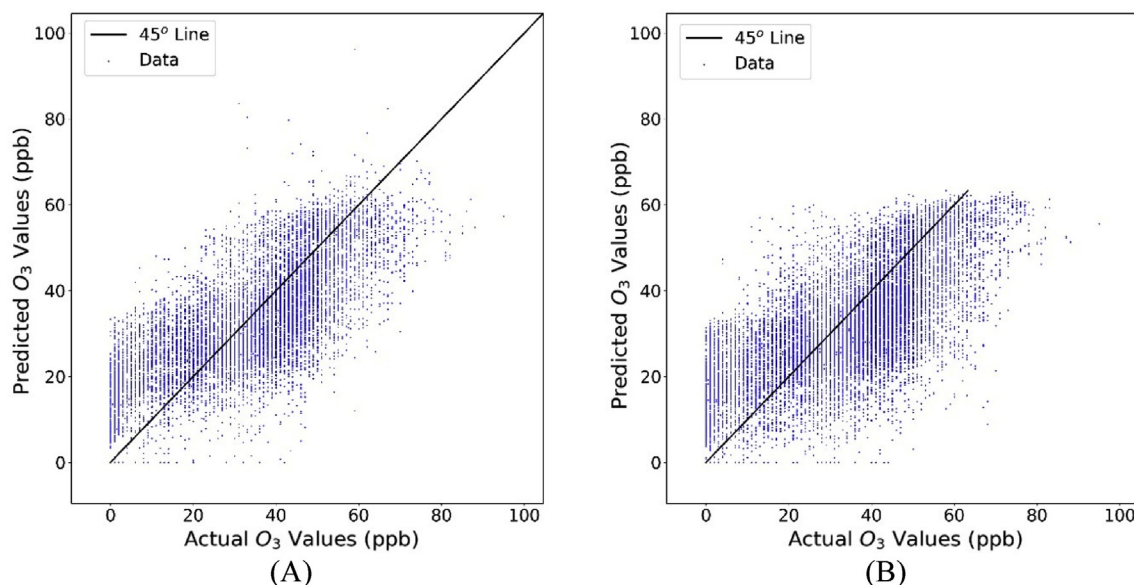


Fig. 3. Parity plots of the test (A) and holdout datasets for MLR model using the four selected variables. The obtained MLR model is as follows: $O_{3,DAQ} = 17.0 - 1.24time + 0.719T_{AirU} + 4.78 \times 10^{-4}OX_{AirU} + 0.013SR_{AirU}$.

3.4.2. ANN model

In order to implement the ANN model to predict the O_3 levels, we applied the optimization again for only four input variables (Fig. S-10). The results indicate that the optimum number of hidden layers and nodes are the same as the ones for all 11 input variables determined in Section 3.3 (1 hidden layer and 25 nodes). Fig. S-11 compares the reference monitor measurements of O_3 with the ANN model estimates for the holdout set over the first week of July 2019. Similar to the MLR model, the ANN model also tracks the O_3 daily cycle. Fig. 4 shows the parity plots for the ANN model. For the test set, The ANN calibration model correlated well with the reference O_3 concentrations (R^2 of 0.763 ± 0.00265) and had a high accuracy (RMSE of 0.533 ± 0.003 ppb). Achieving high R^2 with low RMSE indicates that the developed model is able to explain a large

proportion of the variability in the sensor responses (Ferrer-cid et al., 2019). The low RMSEs can also be attributed to the fact that the MSE (Eq. (2), the square of RMSE) is defined as the loss function for the ANN model. The holdout results also indicate a high level of correlation (R^2 of 0.767) and a low estimation error (RMSE of 0.802 ppb) suggesting that the established ANN model is able to estimate the O_3 readings of the sensors that were not associated with developing the model. The high performance of this model may be attributed to the long calibration period (one year) and a relatively large range of ambient O_3 concentrations (0–95 ppb).

Other studies have established higher calibration correlations ($R^2 > 0.8$) using a variety of nonlinear machine learning algorithms, including RF (Zimmerman et al., 2018), ANN (Spinelle et al., 2015) GPR (Malings et al., 2019), and SVR (Esposito et al., 2016), to predict

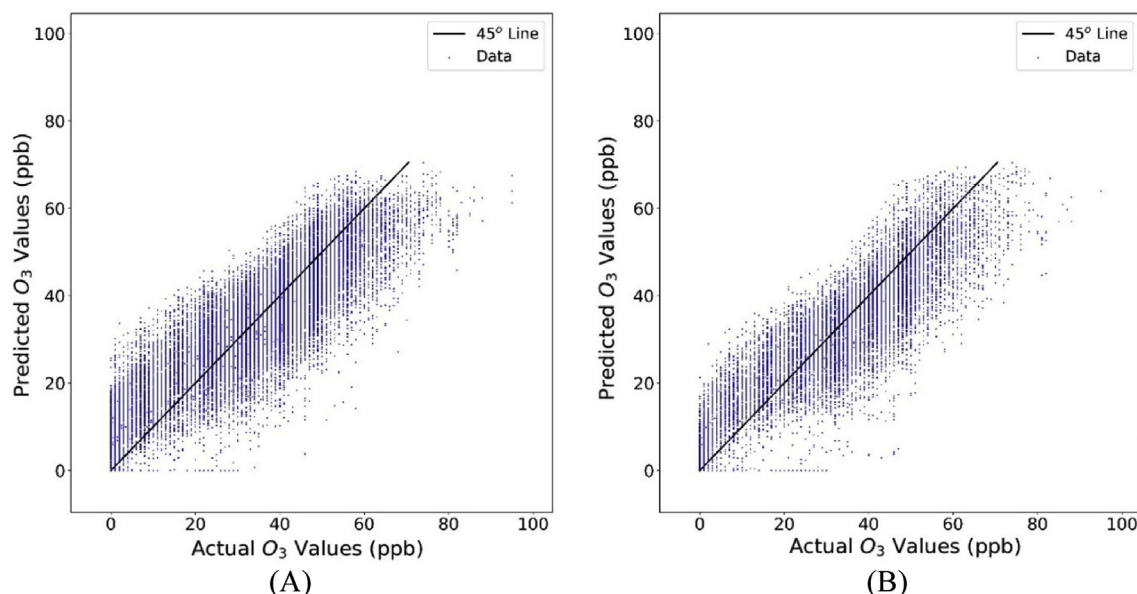


Fig. 4. Parity plots of the test (A) and holdout sets for ANN model using the four selected variables.

ozone concentrations by MO sensors. It should, however, be noted that these calibration models have been established for each individual MO O_3 sensor separately while our model is developed using multiple MO ones (6 sensors) and helps to address intra-sensor variability. Moreover, the calibration time periods of previous studies lasted a few weeks, and the accuracy of their models may deteriorate with time (Ripoll et al., 2019; Spinelle et al., 2015; Zimmerman et al., 2018). The one-year worth of measurement data is anticipated to make the calibration model more robust over time. While this model does not generate sensor-specific responses, it can be applied to other sensors of the same type to produce reasonable estimates of O_3 concentration in an air quality network, such as AQ&U, without having to perform the cumbersome task of individual sensor calibration.

3.4.3. Bias and precision

This study compares the performance of the MLR and ANN models with EPA performance criteria for air quality sensor technologies (Williams et al., 2014). Our MLR model shows relatively high B (41.6%) and P (33.1%). These results indicate that this MLR model is within EPA performance criteria for Tier I purposes (the education and information: $B < 50\%$ and $P < 50\%$). Nevertheless, the ANN model exhibits lower mean B (20.8%) and P (28.6%). The lower mean bias for the ANN model reflects that the difference between the model predictions and the reference monitor readings is small, and the model overestimates the O_3 levels. The small P also demonstrates the ability of the model in predicting the O_3 concentrations. Based on these results, the ANN model meets the EPA criteria for Tier I, Tier II (hotspot identification and characterization: $B < 30\%$ and $P < 30\%$), and Tier IV (personal exposure monitoring: $B < 30\%$ and $P < 30\%$). It should be noted that the EPA does not currently have any defined performance criteria of low-cost sensors for supplementary network monitoring (Tier III).

3.4.4. Post-calibration evaluation

We used the MLR and ANN models to estimate O_3 concentrations over a time period of five months and compared these results with the reference instruments' measurements (Fig. 5). Both the MLR and ANN showed moderate correlations (R^2 s of 0.427 and 0.567, respectively) with high accuracies (RMSEs of 9.99 ppb and

0.635 ppb, respectively) in estimating the O_3 concentrations although the ANN model exhibited better performance. The moderate correlations could be attributed to the low O_3 concentrations (0–49.5 ppb). During the post calibration period, 35.6% of the measurements were lower than the limit of detection (LOD) of the reference monitor (5 ppb, EPA, 2020). During the same five-months calibration period, 24.3% of the measurements were lower than LOD of the reference monitor. Another possible reason could be the lack of O_3 measurements from the Rose Park station (missing 103 days during the same five months of the calibration period). DAQ temporarily closed the Rose Park station from November 18th, 2019 to March 11th, 2019. Additional evaluation of the models' performance at higher O_3 levels is needed to fully understand their performance.

3.4.5. Limitations

The study has some limitations. With six sensors used for training, we might have not captured all the variability of the MO sensors in the network (130 sensors). Moreover, we did not evaluate the effect of the number of sensors used for training and testing on the performance of the calibration models in predicting the holdout set. We only assessed the models' performance for the two holdout sensors during the calibration period and for all eight sensors during five months after calibration period which had a low range of O_3 concentrations. In the future work, we will evaluate the model for a time period after the calibration period when we experience elevated O_3 concentrations, i.e., during warmer summer months. In order to maintain higher quality O_3 concentrations, it would be beneficial to couple the ANN and MLR calibration models with advanced outlier detection (Chen et al., 2018; Ottosen and Kumar, 2019), and/or in-situ calibration methods (Delaine et al., 2019).

4. Conclusions

The primary focus of this study was to develop long-term and generic calibration models that address intra-sensor variability and enhance the quality of O_3 measurements. We implemented a rigorous technique to select physically reasonable input variables using LASSO, ANN and MLR models. We compared a linear (MLR)

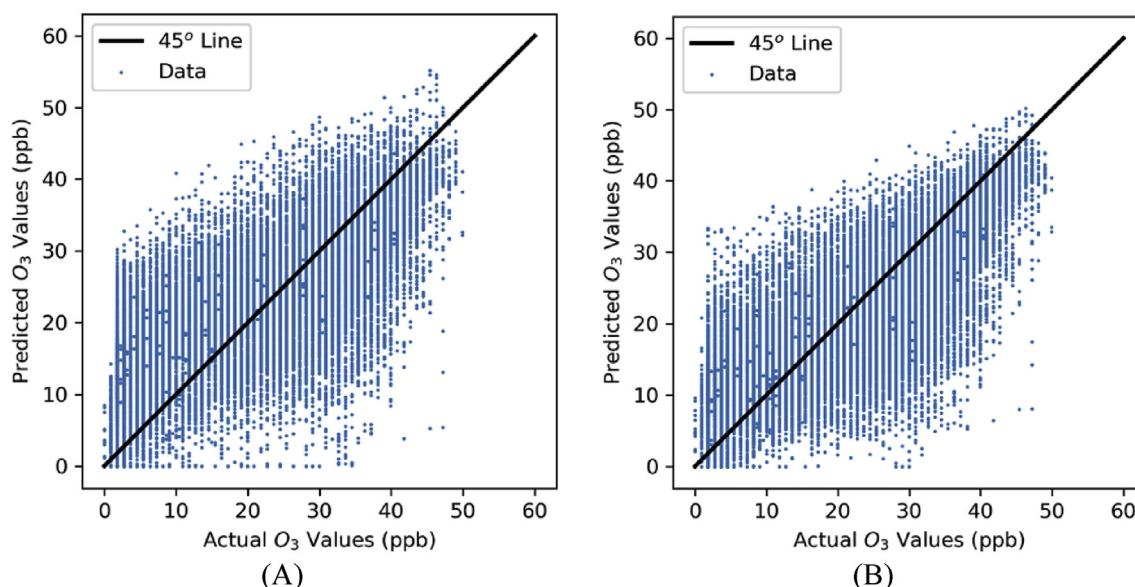


Fig. 5. Parity plots of the MLR (A) and ANN models for five months after the calibration period using the four selected variables.

and non-linear (ANN) calibration models. Both models are able to estimate the O₃ concentrations of other sensors of the same type that were not involved in developing the models. Based on the EPA criteria, the MLR model meets the recommended metrics for Tier I purposes while the ANN model, which exhibited better performance in estimating the O₃ concentrations, is suitable for Tier I, II and IV purposes. The models also showed moderate correlations and high accuracies in predicting the O₃ concentrations for five months after the calibration period. Overall, these calibration models can complement existing regulatory networks to significantly enhance the community-level estimates of O₃ concentration, and help communities and policymakers establishing exposure mitigation strategies.

Author contributions

Tofigh Sayahi: Conceptualization, Formal analysis, Investigation, Methodology, Roles/Writing - original draft; Writing - review & editing. Alicia Garff: Conceptualization, Investigation, Formal analysis, Methodology, Writing - review & editing. Timothy Quah: Formal analysis, Methodology, Writing - review & editing. Katrina Le: Conceptualization, Investigation, Formal analysis, Methodology. Thomas Becnel: Conceptualization, Writing - review & editing. Kody Powell: Conceptualization, Methodology, Writing - review & editing, Supervision. Pierre-Emmanuel Gaillardon: Funding acquisition, Writing - review & editing, Supervision. Anthony Butterfield: Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing - review & editing, Supervision. Kerry Kelly: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Roles/Writing - original draft; Writing - review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Drs. Gaillardon and Kelly, co-authors on this paper, have a financial interest in the company Tetrad: Sensor Network Solutions, LLC, which commercializes solutions for environmental monitoring. Several authors of this paper (Sayahi, Butterfield, Quah, Garff, Becnel, Le, Gaillardon, Kelly) have submitted an invention disclosure related to this paper to the University of Utah technology venture and commercialization office.

Acknowledgement

We gratefully acknowledge National Science Foundation (NSF) support under award numbers 1646408 and 1642513. We are also grateful for support from the Lawrence T. and Janet T. Dee Foundation and UCAIR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2020.115363>.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A System for Large-Scale Machine Learning.

Anenberg, S.C., Horowitz, L.W., Tong, D.Q., West, J.J., 2010. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature

human mortality using atmospheric modeling. *Environ. Health Perspect.* 118, 1189–1195. <https://doi.org/10.1289/ehp.0901220>.

AQ&U, 2019. Air Quality in Salt Lake City [WWW Document]. URL: <http://www.aqanduo.org/>, 5.29.19.

Barcelo-ordinas, J.M., Ferrer-cid, P., Garcia-vidal, J., Ripoll, A., Viana, M., 2019. Distributed Multi-Scale Calibration of Low-Cost Ozone Sensors in Wireless Sensor Networks 1–25. <https://doi.org/10.3390/s19112503>.

Becker, T., Tomasi, L., Bosch-v Braunmühl, C., Müller, G., Sberveglieri, G., Faghi, G., Comini, E., 1999. Ozone detection using low-power-consumption metal-oxide gas sensors. *Sens. Actuators A Phys.* 74, 229–232. [https://doi.org/10.1016/S0924-4247\(98\)00301-X](https://doi.org/10.1016/S0924-4247(98)00301-X).

Becnel, T., Sayahi, T., Kelly, K., Gaillardon, P., 2019a. A Recursive Approach to Partially Blind Calibration of a Pollution Sensor Network. The 15th IEEE International Conference on Embedded Software and Systems (Las Vegas).

Becnel, T., Tingey, K., Whitaker, J., Sayahi, T., Le, K., Goffin, P., Butterfield, A., Kelly, K., Gaillardon, P.E., 2019b. A distributed low-cost pollution monitoring platform. *IEEE Internet Things J.* 6, 10738–10748. <https://doi.org/10.1109/JIOT.2019.2941374>.

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ. Int.* 99, 293–302. <https://doi.org/10.1016/j.envint.2016.12.007>.

Chen, L.J., Ho, Y.H., Hsieh, H.H., Huang, S.T., Lee, H.C., Mahajan, S., 2018. ADF: an anomaly detection framework for large-scale PM2.5 sensing systems. *IEEE Internet Things J.* 5, 559–570. <https://doi.org/10.1109/JIOT.2017.2766085>.

Collier-Oxandale, A., Feenstra, B., Papapostolou, V., Zhang, H., Kuang, M., Der Boghossian, B., Polidori, A., 2020. Field and laboratory performance evaluations of 28 gas-phase air quality sensors by the AQ-SPEC program. *Atmos. Environ.* 220, 117092. <https://doi.org/10.1016/j.atmosenv.2019.117092>.

Cooper, O.R., Gilge, S., Shindell, D.T., 2019. Global Distribution and Trends of Tropospheric Ozone : an Observation-Based Review 1–28. <https://doi.org/10.12952/journal.elementa.000029>.

Delaine, F., Lebental, B., Rivano, H., 2019. In situ calibration algorithms for environmental sensor networks: a review. *IEEE Sensor. J.* 19, 5968–5978. <https://doi.org/10.1109/JSEN.2019.2910317>.

EPA, 2020. AirData Website File Download Page [WWW Document]. URL: https://aqs.epa.gov/aqsweb/documents/data_api.html, 4.21.2020.

Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R.L., Popoola, O., 2016. Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensor. Actuator. B Chem.* 231, 701–713. <https://doi.org/10.1016/J.SNB.2016.03.038>.

Ferrer-cid, P., Barcelo-ordinas, J.M., Garcia-vidal, J., Ripoll, A., Viana, M., 2019. A Comparative Study of Calibration Methods for Low-Cost Ozone Sensors in IoT Platforms. *IEEE Internet Things J.*, p. 1. <https://doi.org/10.1109/JIOT.2019.2929594>.

Goldberger, A.S., 1964. Classical linear regression. In: *Econometric Theory*. John-Wiley & Sons, New York, p. 158.

Gomri, S., Seguin, J.L., Aguir, K., 2005. Modeling on oxygen chemisorption-induced noise in metallic oxide gas sensors. *Sensor. Actuator. B Chem.* 107, 722–729. <https://doi.org/10.1016/j.snb.2004.12.003>.

Gomri, S., Seguin, J.L., Guerin, J., Aguir, K., 2006. Adsorption-desorption noise in gas sensors: modelling using Langmuir and Wolkenstein models for adsorption. *Sensor. Actuator. B Chem.* 114, 451–459. <https://doi.org/10.1016/j.snb.2005.05.033>.

Hagan, D.H., Isaacman-VanWertz, G., Franklin, J.P., Wallace, L.M.M., Kocar, B.D., Heald, C.L., Kroll, J.H., 2018. Calibration and Assessment of Electrochemical Air Quality Sensors by Co-location with Regulatory-Grade Instruments.

Hagler, G.S.W., Williams, R., Papapostolou, V., Polidori, A., 2018. Air quality sensors and data adjustment algorithms: when is it No longer a measurement? *Environ. Sci. Technol.* 52, 5530–5531. <https://doi.org/10.1021/acs.est.8b01826>.

Hitchman, M.L., Cade, N.J., Kim Gibbs, T., Hedley, N.J.M., 1997. Study of the factors affecting mass transport in electrochemical gas sensors. *Analyst* 122, 1411–1417. <https://doi.org/10.1039/a703644b>.

Hsieh, W.W., 2009. *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge university press, Vancouver, BC, Canada.

Hua, Z., Li, Y., Zeng, Y., Wu, Y., 2018. A theoretical investigation of the power-law response of metal oxide semiconductor gas sensors I: Schottky barrier control. *Sensor. Actuator. B Chem.* <https://doi.org/10.1016/j.snb.2017.08.206>.

Jackman, C.H., McPeters, R.D., Labow, G.J., Fleming, E.L., Praderas, C.J., Russell, J.M., 2001. Northern hemisphere atmospheric effects due to the July 2000 solar proton event. *Geophys. Res. Lett.* 28, 2883–2886. <https://doi.org/10.1029/2001GL013221>.

Kelly, K.E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., Butterfield, A., 2017. Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environ. Pollut.* 221, 491–500. <https://doi.org/10.1016/j.envpol.2016.12.039>.

Kingma, D.P., Ba, J.L., 2015. In: Adam: a method for stochastic optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15.

Kouvarakis, G., 2002. Spatial and temporal variability of tropospheric ozone (O₃) in the boundary layer above the Aegean Sea (eastern Mediterranean). *J. Geophys. Res.* 107, 8137. <https://doi.org/10.1029/2000JD000081>.

Lung, S.C., Jones, R., Zellweger, C., Karpinen, A., Penza, M., Dye, T., Hüglin, C., Ning, Z., Lewis, A.C., Schneidmesser, E. von, Peltier, R.E., Leigh, R., Hagan, D., Laurent, O., Carmichael, G., 2018. Low-cost Sensors for the Measurement of

- Atmospheric Composition: Overview of Topic and Future Applications. <https://doi.org/10.1016/j.biopsychem.2014.07.012>.
- Malings, C., Tanzer, R., Haurlyuk, A., Kumar, S.P.N., Zimmerman, N., Kara, L.B., Presto, A.A., Subramanian, R., 2019. Development of a General Calibration Model and Long-Term Performance Evaluation of Low-Cost Sensors for Air Pollutant Gas Monitoring, pp. 903–920.
- Masson, N., Piedrahita, R., Hannigan, M., 2015. Sensors and Actuators B : chemical Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring. *Sensor. Actuator. B Chem.* 208, 339–345. <https://doi.org/10.1016/j.snb.2014.11.032>.
- Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., Mcleod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost , high-density networks. *Atmos. Environ.* 70, 186–203. <https://doi.org/10.1016/j.atmosenv.2012.11.060>.
- Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D.M., Fishbain, B., 2015. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Sci. Total Environ.* 502, 537–547. <https://doi.org/10.1016/j.scitotenv.2014.09.059>.
- Mueller, M., Meyer, J., Hueglin, C., 2017. Design of an Ozone and Nitrogen Dioxide Sensor Unit and its Long-Term Operation within a Sensor Network in the City of Zurich, pp. 3783–3799.
- Ottosen, T.B., Kumar, P., 2019. Outlier detection and gap filling methodologies for low-cost air quality measurements. *Environ. Sci. Process. Impacts* 21, 701–713. <https://doi.org/10.1039/c8em00593a>.
- Park, Y.M., Kwan, M.P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health Place* 43, 85–94. <https://doi.org/10.1016/j.healthplace.2016.10.002>.
- Peterson, P.J.D., Aujla, A., Grant, K.H., Brundle, A.G., Thompson, M.R., Hey, J., Vande, Leigh, R.J., 2017. Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments, pp. 1–25. <https://doi.org/10.3390/s17071653>.
- Ripoll, A., Viana, M., Padrosa, M., Querol, X., Minutolo, A., Hou, K.M., Barcelo-ordinas, J.M., Garcia-vidal, J., 2019. Science of the Total Environment Testing the performance of sensors for ozone pollution monitoring in a citizen science approach. *Sci. Total Environ.* 651, 1166–1179. <https://doi.org/10.1016/j.scitotenv.2018.09.257>.
- Sayahi, T., Butterfield, A., Kelly, K.E., 2019a. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* 245 <https://doi.org/10.1016/j.envpol.2018.11.065> in revision.
- Sayahi, T., Kaufman, D., Becnel, T., Kaur, K., Butterfield, A.E., Collingwood, S., Zhang, Y., Gaillardon, P.-E., Kelly, K.E., 2019b. Development of a calibration chamber to evaluate the performance of low-cost particulate matter sensors. *Environ. Pollut.* 255, 113131. <https://doi.org/10.1016/j.envpol.2019.113131>.
- Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environ. Int.* 106, 234–247. <https://doi.org/10.1016/j.envint.2017.05.005>.
- Sohn, J.H., Atzeni, M., Zeller, L., Pioggia, G., 2008. Characterisation of humidity dependence of a metal oxide semiconductor sensor array using partial least squares. *Sensor. Actuator. B Chem.* 131, 230–235. <https://doi.org/10.1016/j.snb.2007.11.009>.
- Solberg, S., Hov, S., Sørve, A., Isaksen, I.S.A., Coddeville, P., De Backer, H., Forster, C., Orsolini, Y., Uhse, K., 2008. European surface ozone in the extreme summer 2003. *J. Geophys. Res. Atmos.* 113, 9003–9038. <https://doi.org/10.1029/2007JD009098>.
- Sousa, S.I.V., Martins, F.G., 2013. Chemosphere Health effects of ozone focusing on childhood asthma : what is now known – a review from an epidemiological point of view. *Chemosphere* 90, 2051–2058. <https://doi.org/10.1016/j.chemosphere.2012.10.063>.
- Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: ozone and nitrogen dioxide. *Sensor. Actuator. B Chem.* 215, 249–257. <https://doi.org/10.1016/j.snb.2015.03.031>.
- Tibshirani, R., 1996. Regression selection and shrinkage via the lasso. *J. R. Stat. Soc. Ser. B* 1, 267–288.
- Tzortziou, M., Herman, J.R., Cede, A., Loughner, C.P., Abuhassan, N., Naik, S., 2015. Spatial and temporal variability of ozone and nitrogen dioxide over a major urban estuarine ecosystem. *J. Atmos. Chem.* 72, 287–309. <https://doi.org/10.1007/s10874-013-9255-8>.
- Vito, S., De, Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., Francia, G., Di, 2018. Sensors and Actuators B : chemical Calibrating chemical multisensory devices for real world applications : an in-depth comparison of quantitative machine learning approaches. *Sensor. Actuator. B Chem.* 255, 1191–1210. <https://doi.org/10.1016/j.snb.2017.07.155>.
- Williams, R., Kilaru, V.J., Snyder, E.G., Kaufman, A., Dye, T., Rutter, A., Russell, A., Hafner, H., 2014. Air Sensor Guidebook. *Epa/600/R-14/159* 1–5. <https://doi.org/10.1017/CBO9781107415324.004>.
- Zimmerman, N., Presto, A.A., Kumar, S.P.N., Gu, J., Haurlyuk, A., Robinson, E.S., Robinson, A.L., 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* 11, 291–313. <https://doi.org/10.5194/amt-11-291-2018>.