Analyzing and Optimizing an Array of Low-Cost Gas Sensors for use in an Air Quality Measurement Device with Machine Learning

by

Kristen Karl Herod

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Department of Chemical Engineering and Applied Chemistry
University of Toronto

© Copyright by Kristen Karl Herod 2018
Analyzing and Optimizing an Array of Low-Cost Gas Sensors for use in an Air Quality Measurement Device with Machine Learning

Kristen Karl Herod

Master of Applied Science
Department of Chemical Engineering and Applied Chemistry
University of Toronto
2018

Abstract

Low-cost gas sensors have been proposed in place of conventional expensive instruments however they have issues due to cross-sensitivity with other pollutants. Several different types of metal oxide and electrochemical sensors and machine learning methods were evaluated. The objectives were to determine which type of sensor, metal oxide or electrochemical, is better at measuring traffic-related air pollution and whether deep neural networks (DNN) and recurrent neural networks (RNN) improve sensor performance. Three devices were deployed across three sites, two in Toronto and one in Beijing to evaluate the performance of calibration. Calibration was performed with two weeks of data from only one site and evaluated with the remaining data. The combination of metal oxide and electrochemical sensors were more accurate when measuring NO\textsubscript{x}. When targets were normalized, the RNN performed better than DNN and linear calibration, however, not when applied to measuring data well outside the range for calibration.
Acknowledgments

I am very grateful to my Dr. Greg Evans for giving me this research opportunity and providing me with guidance and support. I would like to thank AUG signals for providing funding and for helping develop the AirSENCE device. I would also like to thank the entire SOCAAR team for all their advice which helped me complete my degree.

This research was supported through the Ontario-China Research and Innovation Fund Program. Instruments within SOCAAR were provided through support from the Canada Foundation for Innovation. I am grateful to Jerzy Deboz and other staff at the Ontario Ministry of the Environment and Climate Change who operated and provided access to data from the monitoring station at the Highway Site. This involvement by the MOECC does not imply endorsement or support for the commercialization of any particular sensor based air quality technology.
# Table of Contents

Acknowledgments ........................................................................................................................................ iii

Table of Contents ...................................................................................................................................... iv

List of Figures ......................................................................................................................................... viii

Chapter 1 ................................................................................................................................................... 1

1 Introduction .......................................................................................................................................... 1

1.1 Context and problem definition ...................................................................................................... 1

1.1.1 Types of low-cost sensors ........................................................................................................ 1

1.1.2 Types of calibration models .................................................................................................... 2

1.2 Overview of Research ................................................................................................................... 2

1.2.1 Research Questions .................................................................................................................. 2

1.2.2 Summary of thesis sections .................................................................................................... 3

Chapter 2 ................................................................................................................................................... 4

2 Literature review ................................................................................................................................... 4

2.1 Health effects of traffic related air pollution ................................................................................ 4

2.2 Low cost gas sensors ................................................................................................................... 5

2.2.1 Introduction to low cost gas sensors ...................................................................................... 5

2.2.2 Calibration of metal oxide sensors ......................................................................................... 7

2.2.3 Calibration of electrochemical sensors .................................................................................. 9

2.2.4 Effect of drift on low-cost sensors ....................................................................................... 11
2.3 Calibration with non-linear models ................................................................. 11
2.4 Evaluation of current sensors ........................................................................ 14
2.5 Conclusions ...................................................................................................... 14

Chapter 3 ................................................................................................................. 16

3 Methods ................................................................................................................ 16

3.1 Device Design .................................................................................................. 16

3.1.1 AirSENCE device and sensors .................................................................... 16

3.2 Device Calibration ............................................................................................ 17

3.2.1 Reference instruments .................................................................................. 17

3.2.2 Study sites and devices ................................................................................ 17

3.3 Data Analysis .................................................................................................. 19

3.3.1 Cross-responsiveness with other gases and environmental factors .......... 19

3.3.2 Feature selection ........................................................................................... 21

3.3.3 Normalization ............................................................................................... 23

Chapter 4 ................................................................................................................ 25

4 Sensor based measurements of Nitrogen oxides using three array designs ........ 25

4.1 Calibration and evaluation ................................................................................ 25

4.2 Metal oxide, electrochemical, and combined sensor array performance at highway site 25

4.3 Site selection for sensor training ...................................................................... 29

4.4 Training to application site transferability ......................................................... 31
Chapter 5 ....................................................................................................................................... 34

5  Comparison of different models used for calibration of low-cost sensors  .................. 34

5.1 Deep neural network ...................................................................................................... 34

5.2 Recurrent neural network and Long Short-Term Memory ............................................ 35

5.3 Comparison of MLR, DNN, and RNN for measuring air pollution using low-cost sensor array 36

5.4 Comparison of different training sites for calibrating low-cost sensor array ............... 42

5.5 Effect of normalizing reference data for training models ............................................. 44

5.6 Model performance with cross-sensitivity ..................................................................... 46

5.7 Model performance at different training sites .............................................................. 47

Chapter 6 ....................................................................................................................................... 51

6  Applications of low-cost air quality devices......................................................................... 51

6.1 Highway gradient study ................................................................................................. 51

6.1.1 NO results ............................................................................................................. 51

6.1.2 NO₂ results ......................................................................................................... 54

6.1.3 O₃ results .............................................................................................................. 57

6.2 Train study .................................................................................................................... 59

6.2.1 NO results ............................................................................................................. 60

6.2.2 NO₂ results ......................................................................................................... 61

Chapter 7 ....................................................................................................................................... 65

7  Conclusion and future directions ...................................................................................... 65
7.1 Conclusion........................................................................................................................................... 65

7.2 Future directions................................................................................................................................... 66

References.................................................................................................................................................. 68
List of Figures

Figure 1: Metal oxide sensing principles [41] ................................................................. 6

Figure 2: Electrochemical sensing principles [23] ......................................................... 7

Figure 3: Correlation matrix of all sensors ................................................................. 22

Figure 4: Box and whisker plot of NOx concentration measured June 20 – July 6, 2016 at the Highway Site. The edges of the box represent the 25th and 75th percentiles, the center line is the median, the whiskers are the maximum and minimum, and the dots are the outliers. The maximum and minimum are defined as 1.5 times the interquartile range above the 3rd quartile and below the 1st quartile and outliers are outside that range. Each of the groups of boxes represents increasing 50 ppb levels of reference NOx data. The electrochemical is closer to the reference measurements than metal oxide sensors, more notably at higher concentrations........ 26

Figure 5: NOx residuals (reference minus sensor) of NOx for metal oxide sensor array show significant cross sensitivity with O3 for NOx concentration below 50 ppb (left) however above 50 ppb NOx they do not (right) .................................................................................................... 27

Figure 6: Box and whisker plot of NOx concentration measured June 20 – July 6, 2016 at the Highway Site. The edges of the box represent the 25th and 75th percentiles, the center line is the median, the whiskers are the maximum and minimum, and the dots are the outliers. The electrochemical sensors alone yielded data with similar distribution to that from a combined array of metal oxide and electrochemical sensor types .......................................................... 28

Figure 7: The combined array was calibrated at the Downtown and Highway Sites and then tested against reference data from both sites (slopes = 0.6 for Downtown and 0.9 for Highway). Shaded area represents 95% confidence interval. The dashed line is for a perfect 1:1 model. .... 30

Figure 8: Box and whisker plot for sensors calibrated at the Highway Site and then used for NOx measurements at the Highway or Downtown Sites. The edges of the box represent the 25th and 75th percentiles, the center line is the median, the whiskers are the maximum and minimum, and
the dots are the outliers Spatial difference between the distributions Downtown and at Highway are higher with electrochemical sensors. ................................................................. 31

Figure 9: Performance trade-off of the metal oxide sensors, electrochemical and combined array designs in the measurement of NOx. Relative error is calculated as the average of the reference values minus the predicted values divided by the reference values................................. 33

Figure 10: Comparison of DNN, MLR, and RNN for testing data from all devices calibrated using two weeks of data from Device 1 (Toronto) at the Downtown Site. The line represents the median and the shaded area represents the 10th and 90th percentile of the data......................... 42

Figure 11: Comparison of DNN, MLR, and RNN for all testing data from all devices calibrated using two weeks of data from Device 1 (Toronto) at the Highway site. The line represents the median and the shaded area represents the 10th and 90th percentile of the data......................... 44

Figure 12: Comparison of RNN with normalized target variables and unnormalized target ...... 45

Figure 13: Comparison of RNN with normalized and unnormalized target at all sites trained using two weeks of data from Device 1 (Toronto) at the Downtown site. Device 1 (Toronto) was tested at both the Downtown and Highway site, Device (PKU) was tested at Peking University, Device 2 was tested at the Downtown site, and Device 3 was tested at the Highway site........... 46

Figure 14: Distribution of reference NO for calibration data Downtown and at the Highway compared to the distribution of all data from Device 1. ................................................................. 48

Figure 15: Distribution of reference O3 for calibration data Downtown and at the Highway compared to the distribution of all data from Device 1. ................................................................. 49

Figure 16: Distribution of reference NO2 for calibration data Downtown and at the Highway compared to the distribution of all data from Device 1. ................................................................. 50

Figure 17: Hourly AirSENCE NO measurements next to Highway 401 and 200 meters away from Highway 401 show variation between the two sites.............................. 52

Figure 18: Boxplot comparing Reference and AirSENCE NO measurements next to Highway 401 and 200 meters away from Highway 401. ................................................................. 53
Figure 19: Linear relationship in the difference in NO concentrations next to highway and 200 meters away from highway.............................................................................................................. 54

Figure 20: Time series of AirSENCE NO2 measurements next to Highway 401 and 200 meters away from Highway 401.............................................................................................................. 55

Figure 21: Boxplot comparing Reference and AirSENCE NO₂ measurements next to Highway 401 and 200 meters away from Highway 401. ................................................................. 56

Figure 22: Boxplot comparing Reference and AirSENCE NOx measurements next to Highway 401 and 200 meters away from Highway 401. .............................................................................. 57

Figure 23: Timeseries of AirSENCE O3 measurements next to Highway 401 and 200 meters away from Highway 401.............................................................................................................. 58

Figure 24: Boxplot comparing Reference and AirSENCE NO₂ measurements next to Highway 401 and 200 meters away from Highway 401. .............................................................................. 59

Figure 25: Hourly reference NO measurements vs. low cost sensors calibrated in laboratory and using RNN during commuter train study................................................................. 61

Figure 26: Reference NO₂ measurements vs. low cost sensors calibrated using RNN during commuter train study. ............................................................................................................... 62

Figure 27: Reference NO₂ measurements vs. low cost sensors calibrated using RNN without SN706 sensor. Overview of low cost sensor application......................................................... 63
Chapter 1

1 Introduction

1.1 Context and problem definition

1.1.1 Types of low-cost sensors

In urban ambient settings, traffic can be a major source of pollution and exposure to these emissions has been associated with a range of adverse health outcomes [1–7]. Traffic related air pollution concentrations can vary spatially in urban areas due to varying proximity to roadways [8,9]. High spatial resolution measurement is therefore required to accurately assess the resulting exposure. Nitrogen oxide (NO) nitrogen dioxide (NO₂), and carbon monoxide (CO) are major pollutants in traffic emissions. Ozone (O₃) is a secondary pollutant whose concentrations can also be influences by traffic emissions. In near-road region, NO is typically directly emitted by vehicles, but this is rapidly converted to NO₂ by reaction with O₃. Conversely, NO₂ can be photochemically converted back to NO at a rate that depends on solar intensity. Due to this rapid interconversion, the ratio of NO to NO₂ varies in both time and space. Thus, measurement of the sum (\( \text{NO}_x = \text{NO} + \text{NO}_2 \)) is often simpler and more useful, particularly when examining traffic related gradients across cities.

Low-cost (< $10,000) gas sensors provide a scalable and accessible alternative to the larger and more expensive instruments used to measure gaseous pollutants for regulatory purposes. Regulatory-grade instruments are difficult to transport and too expensive for most applications outside of governmental monitoring environments [10,11]. Although low cost devices are unlikely to achieve as high an accuracy and reliability as their expensive counterparts, they may still open the door to broad and accessible coverage of air quality measurements across cities. A key element to realizing this vision of creating networks across smart cities [12] in the future is demonstrating that sensor-based air quality monitoring devices can reliably measure spatial heterogeneity of ambient pollutants within a city [13]. Low cost sensors have been applied to quantifying ambient concentrations of NO₂ [14–18], NO [19], and both [20–23]. However, few studies have investigated using low cost sensors to measure total \( \text{NO}_x \) directly.
1.1.2 Types of calibration models

Calibration is required to interpret the values from a low-cost gas sensor since changes in concentration are measured as changes in voltage. Calibration can be done with many different kinds of models like multiple linear regression (MLR) and non-linear models such as neural networks. Neural networks were initially used with low cost sensors mainly for gas identification [24–27]. In more recent years non-linear models have been applied to predicting gas concentration values using arrays of low cost sensors which is a regression problem. Several papers have used feed forward neural networks in order to calibrate low cost sensor devices using ambient data and have obtained better results than linear models [17,28]. In another study, random forest models were used to calibrate sensors and yielded better results than linear models [29]. In one study feed forward neural networks, tapped delay neural networks, and nonlinear autoregressive with exogenous inputs networks were compared to each other [30]. In that study only electrochemical sensors were used for gas measurement and temperature, relative humidity, wind direction, and wind speed were used as input features for the model.

1.2 Overview of Research

1.2.1 Research Questions

There are many challenges related to development of low-cost devices for measuring NOₓ and O₃. One unresolved question is which type of sensor, metal oxide, or electrochemical is better for measuring pollution and whether combining these sensors improves the accuracy. There has been much research done on calibrating devices in laboratories with standard gases and some research done calibrating devices in ambient conditions however there has not been research into how much of a difference it makes where these devices are calibrated. Low cost devices generally contain an array of low cost gas sensors and environmental sensors. In this thesis non-linear models are explored to see how they compare to each other and linear models. Finally, rigorous evaluation techniques, such as using several independent sites are used to assess the accuracy and general readiness of these types of low-cost devices to be used for applications such as measuring spatial variation and indoor air quality.
1.2.2 Summary of thesis sections

In the second chapter, a synthesis of literature review is provided to establish the gaps in the field. First, the health effects associated with air pollution, and specifically traffic related air pollution are explored. The spatial differences associated with urban air pollution are reviewed to establish the need for high spatial resolution air quality monitoring networks. Then, the past research done into calibration of metal oxide and electrochemical type low-cost sensors are explored.

The methods are then described including the low-cost sensors and reference instruments used, along with the sites used for calibration and evaluation. An overview of the challenges with low-cost sensors such as drift and cross-sensitivity are also explored in depth. Finally, the analysis techniques such as feature extraction, normalization, calibration methods, and evaluation methods used are explored.

In Chapter 4, results from the first major deployment are presented. In this sub-study, three different sensor arrays were investigated for measuring NO$_x$ including only metal oxide sensors, only electrochemical sensors, and a combination of both sensor types. Then MLR models were calibrated at different sites to assess how much of a difference choosing a site makes. Finally, these sensor arrays were evaluated at a site which was not used for calibration to assess how well the calibration models transfer between sites. Also, the comparison of price vs. accuracy was examined to establish which combinations of sensors best meet the criteria for low cost devices.

In Chapter 5 non-linear calibration models are examined. These calibration models were trained using one device and applied to all devices to explore whether one model can be used accurately for multiple devices. Also, three separate sites with reference data were used to assess how well the models perform when sent to new environments. Finally, the effect of normalizing reference pollutants for optimizing neural network performance was explored.

In Chapter 6, a prototype sensor device, optimized with the methods developed in the earlier chapters was applied to several “real world” health related applications such as determining the spatial variation of pollution concentration next to major roadways and measuring the concentrations emitted in a commuter train. The devices in this chapter used the same models that were calibrated in the previous section.
Chapter 2

2 Literature review

2.1 Health effects of traffic related air pollution

According to the World Health Organization (WHO), air pollution is responsible for 7 million premature deaths annually [31]. In Shanghai, China, Chen et al. showed that the air quality health index (AQHI) was significantly related to daily mortality [32]. Maji et al., showed that elevated fine particulate matter (PM$_{2.5}$) in 161 Chinese cities was associated with over 600 000 premature deaths in 2015 [33].

Traffic emissions can be a major contributor to air pollution and several studies have linked proximity to major roadways with adverse health effects. One such study found that those living within 50 meters of a highway had a 1.56 increased risk of sudden cardiac death compared to those living 500 meters away [2]. Although there is a link to traffic, specific pollutants were not examined, only proximity to the highway. In another study, the effect of living near a major road was examined for incidence of dementia, Parkinson’s disease, and multiple sclerosis. It was found that there was a hazard ratio for dementia of 1.07 for people living within 50 meters from a major road and 1.04 and 1.0 for people living between 50-100 and 100-200 meters away from major roads respectively. However, road proximity did not influence incidence of Parkinson’s disease or multiple sclerosis [5]. Furthermore, other studies have linked increased concentrations of pollutants such as O$_3$ and NO$_2$ to poor health. It was found that for each 24 ppb increase in O$_3$ there was a 2.5 percent increase in systolic blood pressure and a 3.26 percent increase in diastolic blood pressure after 3 hours of exposure among women [3]. In another study, ischemic stroke incidence rate ratios (IRR) of 1.11 were found at 10 µg/m$^3$ of NO$_2$ and 10 db of noise from road traffic [4]. Traffic-related air pollution has also been found to have potential effects on the health and development of children. Cognitive development of primary school children was also examined at schools with different pollution levels. It was found that a change of NO$_2$ and elemental carbon (EC) concentration from 1$^{st}$ to 4$^{th}$ quartile reduced gain in working memory by 13 percent in children [6]. Also, children with the GSTP1 minor allele had an increased risk of asthma (OR = 2.59) per 10 µg/m$^3$ increase in NO$_2$ [7].
Measuring air pollution remains a challenge in urban settings since concentrations may vary greatly between different neighborhoods. NO and NO$_x$ were compared across three different neighborhoods. The average NO and NO$_x$ concentrations for the three neighborhoods, relative to the entire study, were 0.64 and 0.96, 0.50 and 0.75, 0.48 and 0.68 respectively. NO and NO$_x$ vary quite a lot between neighborhoods however NO$_2$ does not vary as much [8]. In another study three sites were also compared in the Boston area. The median NO and NO$_x$ concentrations at these three sites were 15 and 33 ppb, 31 and 67 ppb, 16 and 36 ppb respectively. As well as inter-neighborhood differences, this study also explored the effect of distance from major roadways on pollutant concentration. In the previous section, several health studies linked distance to roadways as having an effect on health. At two sites in Boston, pollution decreased by 34 and 30 percent from 0 to 200 meters away from a highway [9]. Brantley et al. used mobile sampling to measure the difference in pollution between different areas. The pollution along the selected routes ranged from below 10 µm/m$^3$ to over 90 µm/m$^3$ [34]. Kasner et al. measured the gradient of CO$_2$, NO, NO$_2$, O$_3$, and many other pollutants at various distances from a major highway using a mobile sampling device. Downwind of the highway the average rush hour NO$_x$ concentration changed from about 80 ppb to 40 ppb across a distance of 400 meters [35]. A study by Karner et al. showed that NO$_2$, NO$_x$, and O$_3$ all reached approximately background levels at a distance of 570 meters from the edge of a road [36].

In a large urban environment, having only a few monitoring stations may not be enough to account for this inter-neighborhood variation. Conventional research requires the use of large instruments that require constant calibration and are static [37–39]. Therefore, other methods are important to fill in the gaps that conventional monitoring may leave. The focus of this thesis will be on the challenges and advantages of using low-cost sensors to measure air pollution.

### 2.2 Low cost gas sensors

#### 2.2.1 Introduction to low cost gas sensors

There are many different manufacturers of metal oxide sensors and they are all relatively inexpensive ($6 ~ $40 USD). Metal oxide sensors consist of a sensitive layer, a substrate, electrodes, and a heater resister. The sensitive layer is made from a metal oxide such as tin oxide and is typically doped with material that is sensitive to the target gas. When voltage is applied across the heater, the sensor surface becomes very hot (~200 - 300 °C), which makes the
sensitive layer catalytically active. The ambient air then reacts with the surface of the sensitive layer by either adding or removing electrons, which get replaced from the bulk layer where the excess charge is stored known as band bending. As charge is added or removed from the bulk layer, the conductance changes as a function of the target gas (Figure 1). These sensors are not perfectly selective as the sensitive layer can interact with other types of pollutants [40]. For instance, NO\textsubscript{x} sensors are very sensitive to O\textsubscript{3} [12]. Sensor readings can also be affected by other environmental factors such as temperature and humidity, causing the sensors to degrade over time that causes values to “drift” and lose sensitivity to the target gas. As a result, these sensor readings must either be corrected for or replaced after a certain period of time, which can lead to extra costs when applying these for long term monitoring. Metal oxide sensors alone are not very effective for measuring ambient NO\textsubscript{x} concentrations, however if they are combined, their performance can increase considerably.

![Figure 1: Metal oxide sensing principles [41].](image)

Electrochemical sensors are another type of low cost gas sensor that can now be used for ambient pollution detection due to recent advances in the technology [42]. These sensors are marginally more expensive than metal oxide sensors (~$60 USD for Alphasense) and usually require a separate circuit board (~$90 USD for Alphasense). Electrochemical sensors consist of a gas permeable membrane, seal, working electrode, reference electrode, counter electrode, electrolyte reservoir, wetting filters, and in some cases an auxiliary electrode. The basic principles of operation are that the gas enters through the hydrophobic membrane that prevents the electrolyte from escaping the sensor (Figure 2). The gas then reacts at the surface of the sensing electrode by
oxidation or reduction. These reactions are catalyzed at the surface of the electrode by materials that are designed for the target gas. There is a resistor connected across the electrodes in which a current proportional to the gas concentration flows from anode to cathode. The reference is used to measure and correct for the driving voltage since it is not constant at the sensing electrode [42].

![Electrochemical sensing principles](image)

These sensors can be sensitive to other gases such as CO and O₃ as well as environmental factors like temperature [43], a detailed list of sensor sensitivities is shown in Table 1 (pg.16). Temperature can be corrected for by using the auxiliary electrode that measures the influence of temperature on the sensor. These sensors have been found to not exhibit drift over a few months in typically polluted areas [19]. However, these electrochemical sensors are very sensitive to changes in pressure and the manufacturer recommends that they be calibrated and tested as close to the same atmospheric pressure as possible. Low pressures may cause the electrolyte to escape the membrane and high pressures will show higher than normal values. According to the manufacturer of the Alphasense electrochemical sensor, the sensor can operate in up to 20 kPA gauge pressure, however it is still recommended to keep the sensor at atmospheric pressure [44]. This pressure sensitivity can make calibrating sensors using pumped air difficult to apply to ambient settings where the pressure may differ.

### 2.2.2 Calibration of metal oxide sensors

One early application of low-cost sensors was to identify gases based on the response curves of sensors under modulating temperature [45–48]. However, in order to measure gas concentration, these changes in resistance must be calibrated with some model that can transform the values to
the corresponding concentration. There are many methods for calibration of metal oxide sensors including under laboratory conditions and under ambient conditions. Under laboratory conditions “standard gas” cylinders of the target pollutant are introduced at different concentrations and other variables such as temperature and humidity are controlled for. Under ambient conditions sensors are generally placed in close proximity to a reference instrument and the values are trained to match the reference instrument values.

Carotta et al. showed that their NO\textsubscript{x} sensor had a $R^2$ of 0.86 with the reference instrument when evaluated outdoors [49]. The $R^2$ obtained in this study was higher than many other studies where commercially available sensors were used. In another study, metal oxide sensors were calibrated using both laboratory and ambient conditions and evaluated outdoors next to a reference instrument. For measuring CO, the model trained in the laboratory had a median error of 3.56 ppm and the model trained using collocation with a reference instrument had a median error of 0.45 ppm. The lab calibration performed much worse, which is likely because it did not take into account other factors that affect the sensors such as cross-sensitivity with other pollutants [16]. Another study also calibrated a sensor array with both an artificial mixture of pollutants as well as using outdoor pollution measurements. Although they only had a couple of days of measurements for evaluation, “making a calibration model with artificial mixtures of some pollutants cannot be successful” [50]. Wen et al. calibrated a metal oxide CO sensor by placing it next to a reference station and finding the relationship between them. The sensor voltage had a “break point” after which it had a linear relationship with reference CO and a $R^2$ of 0.76 [51].

Based on the literature mentioned in this paragraph, when using metal oxide sensors for measuring ambient pollution concentrations, it is best to coloclate the sensor array next to a reference instrument since the combination of gases in ambient conditions has an effect on the resistance of the sensors.

In order to reduce the affect from cross-sensitivity, Spinelle et al. created a low-cost sensor array using many different commercially available metal oxide sensors and calibrated this array using a linear regression and a multiple linear regression. The multiple linear regression for NO\textsubscript{2} was trained using the sensor resistance, reference O\textsubscript{3}, and temperature. For NO\textsubscript{2}, the highest $R^2$ using a linear regression was 0.20 and using multiple linear regression was 0.06 [17]. This could mean that the multiple linear regression weights were over fitted for temperature or humidity during the calibration period. One drawback with the studies mentioned above is that the devices were
evaluated at the same site as where they were calibrated. Therefore, the sensor array may be over fitted for the specific composition of pollutants in that site. This thesis focuses on how to simulate “real world” conditions and variability by evaluating sensors at a site that was not used for calibration. Another method used to calibrate metal oxide sensors was step calibration. This type of calibration was done by regulating the flow to the sensor and “zeroing” the sensor every so often. This sensor was tested at various reference sites across the world. Some examples of this sensor performance were in Houston, USA, where the standard error was less than 5 ppb and in Auckland, New Zealand where the error was less than 2 ppb [39].

2.2.3 Calibration of electrochemical sensors

Several studies have applied electrochemical sensors for ambient concentration measurements. In one study electrochemical sensors were used to measure NO, CO, and NO₂ Sensors were calibrated using calibration gas in a laboratory setting then evaluated in ambient conditions. NO sensors had R² values of 0.80 and 0.95 for two collocated sensors and 0.89 and 0.92 for NO₂, after NO₂ values were corrected with O₃. The O₃ measurements were only available from the reference instruments, which means that this correction method is not practical unless it is measured separately and accurately using low cost sensors [23]. In another study using only electrochemical sensors, an NO sensor was calibrated in a ambient setting and evaluated outdoors [19]. The electrochemical sensor had an exponential relationship with temperature and experienced higher error above 80% relative humidity. After deriving an equation using sensor values and temperature, the sensor had a RMSE of 14.9 ppb. In a similar study by Popoola et al., NO and CO were measured using only electrochemical sensors and a correction for temperature was developed. The sensors were collocated alongside reference instruments and models were developed with temperature correction. Before correction, the R² of the two NO sensors and the reference instrument was 0.02 and after temperature correction, it was 0.84 and 0.87. The CO values could not be compared at the secondary site since there was no reference CO [52]. Sun et al. developed a similar low-cost device using electrochemical sensors for the Hong Kong Marathon in 2015. Sensors were calibrated in a laboratory setting by introducing known concentrations of standard gas for NO₂ and CO at various temperatures and humidity. In the field
tests, over a 2-day period, the CO sensor had a $R^2$ of 0.97 and the NO$_2$ sensor had a $R^2$ of 0.9 with reference instruments [18].

Pang et al. developed a low cost O$_3$ measuring device using an electrochemical sensor. The sensor was calibrated in a laboratory setting by generating known concentrations of O$_3$ and measuring the electrical response of the sensor. The O$_3$ sensor needed to be corrected for NO$_2$, NO, and relative humidity under ambient conditions since it is cross sensitive to those. In ambient conditions the corrected sensor values had a $R^2$ of 0.84 over an 18-day period and values reached 60 ppb. The sensor over predicted the concentration slightly with a slope of 1.09 compared to the reference monitor [53].

In a study by Spinelle et al. both electrochemical and metal oxide sensors were examined for measuring NO, CO, NO$_2$, CO$_2$, and O$_3$. For NO, only electrochemical sensors were used. With a linear model the highest $R^2$ achieved was 0.02 and using a multiple linear regression 0.02 was also achieved. The type of electrochemical sensor used was not Alphasense, which was used in every other study, but CitySense. For NO$_2$ measurement, the electrochemical sensor, which was Alphasense, had a $R^2$ of 0.01 with a linear regression and 0.09 with a multiple linear regression. The multiple linear regression included temperature, relative humidity, and reference O$_3$. For measuring O$_3$ using an Alphasense electrochemical sensor, the $R^2$ with a linear regression was 0.2 and the $R^2$ with a multiple linear regression was 0.48. The other O$_3$ electrochemical sensor, CitySense, had a $R^2$ of 0.67 using linear regression and 0.8 using multiple linear regression [17,28].

When evaluating the performance of electrochemical sensors there seems to be a wide range of results. From the studies above, the Alphasense NO-B4 electrochemical sensor seems to perform quite well when it is corrected for temperature. The CO-B4 electrochemical sensor also seems to perform well however the temperature correction is not as good as with the NO-B4 sensor. For O$_3$ and NO$_2$ there is a wide variety of different results with some studies achieving satisfactory results with electrochemical sensors and other which do not. Many studies need to correct NO$_2$ with O$_3$ since the two pollutants are cross-sensitive to each other [14,17,23]. This may leave the door open to combining metal oxide and electrochemical sensors together to account for the shortcomings of each individual sensor. From the current literature, electrochemical sensors
seem better for measuring NO and CO however, for NO₂ and O₃ it is not clear which sensor type is better.

2.2.4 Effect of drift on low-cost sensors

Drift remains an important challenge when using low cost air quality sensors especially with metal oxide type sensors. Many studies have been done in order to reduce the effect of drift [41,54]. In a paper by Masson et al., the drift apparent with metal oxide sensors seems to be due to the high heat affecting the heating resistor and making it less effective over time. This means that as the heat degrades the resistor, it becomes less effective at heating and the temperature will decrease over time. As the temperature decreases the surface site becomes less catalytically active and does not react with pollutants. This affects both the baseline value and the sensitivity of the sensor over time. For a CO metal oxide sensor, the baseline resistance of the sensor decreased by approximately 10 kohm after 15 days [41]. Most studies looking into the effects of drift are focused on the metal oxide type of sensors however electrochemical sensors do not seem to be affected by drift for at least the first 6 months [19]. In this thesis drift was accounted for using minutes from deployment as an input to the model and longer training time. This method for drift correction performed worse with the evaluation data so it was not used. The effect of cross-sensitivity to gases and temperature outweighed the effect from drift when using a linear and non-linear models.

2.3 Calibration with non-linear models

In previous sections, metal oxide and electrochemical sensors have been explored for use in low-cost air quality measurement devices. These sensors have been calibrated in a lab setting, using known concentrations of standard gas, and variations in temperature and humidity to derive a model. However, this method of calibration has drawbacks since other gases present in ambient air may have an effect on the sensor readings. As a result, arrays of sensor, designed to measure multiple pollutants were collocated alongside reference instruments outdoors and calibrated using linear models. Since sensors can be affected by temperature, humidity and other gases, multiple linear regressions were used which combine several sensor readings, and even reference concentrations, in order to account for this. Other studies have gone one step further by applying non-linear models in order to calibrated sensors.
A feed forward neural network (FFNN) is the simplest type of neural network, where values get passed through hidden layers and then output. The hidden layers consist of n hidden nodes and have an activation function which can be linear, sigmoid, tan, etc. Before inputs are passed through hidden nodes they are multiplied by weights which are randomly generated and optimized using backward propagation. Backward propagation is a process where outputs are compared to the real target values and then the estimated optimizations to the weights are passed backward through the network [55,56]. Unlike a linear regression, a neural network can handle non-linear problems as well. However, since the weights are randomly initialized you may get different weights every time you train the model unless a seed is used. A seed is used to create the same randomly generated numbers each time. When training neural networks a local minima can also be reached and the model will never reach the optimal solution.

Non-linear models, such as neural networks, have been used for a long time for gas identification with low cost sensors [57–61]. Heilig et al. used one metal oxide sensor to identify CO and NO$_2$ and predict the concentration by using a technique known as temperature modulation coupled with a neural network [24]. The drawback of this study was that only standard gas was used and it was not evaluated under ambient conditions. Sommer et al. also used metal oxide sensors for gas identification using temperature modulation and neural networks. Using this method they were able to distinguish butane and methane with a maximum error of 0.07 and 0.51 vol.% respectively [26]. Qihe Liu et al. and Hang Liu et al. developed methods for metal oxide sensor drift compensation using deep neural network (DNN)$^1$ [27,54].

Recently, non-linear models have been used for calibrating low cost air quality sensors for ambient gas concentration measurement. Zimmerman et al. calibrated arrays of low cost gas sensors for measuring CO, CO$_2$, NO$_2$ and O$_3$ using a Random Forest (RF)$^2$ model. The mean absolute error (MAE) for was 38 ppb for CO, 10 ppm for CO$_2$, 3.5 ppb for NO$_2$, and 3.4 for O$_3$ for the random forest which performed better than lab calibration and multiple linear regression models [29].

In one study low cost sensors were calibrated using neural networks and compared to linear models. For measuring NO$_2$ and O$_3$, the best R$^2$ was 0.60 and 0.92 respectively which was better

---

$^1$ DNN is a non-linear model which is similar to a neural network except there must be more than one layer of hidden nodes

$^2$ RF is a non-linear model which uses multiple randomly created decision trees which “vote” on the best output
than with linear or multiple linear regression models as mentioned in the previous section [17,28]. Another study used a neural network with “fuzzy logic” type pre-treatment to train low cost sensors to identify three types of sources: traffic, urban, and photochemical pollution. This method achieved an accuracy of 80 percent however this method only classifies which type of source but not which type of pollutant there is and what the concentration is [62].

One issue that arises when using non-linear models such as neural networks is that they may over fit the training data and then perform worse under testing conditions. In order to mitigate this, one study investigated the use of Bayesian Regularization in order to reduce the effect of overfitting. Bayesian Regularization uses Bayesian statistics in order to optimize when the FFNN stops learning which reduces the effect of overfitting [63]. Using Bayesian regularization as part of the calibration, the optimal training length used for calibrating models was found to be about two weeks [21].

Some other studies have used tapped delay non-linear models (TDNN)\(^3\), which use past sensor data and current sensor data in order to predict gas concentration using low cost sensors. One study used this method along with neural network (NN) and support vector machine (SVM)\(^4\) in order to predict NH\(_3\), NO\(_2\), and RH. The NN had a mean relative error (MRE) of 7.2, 8.2, and 7.9 percent for RH, NO\(_2\), and NH\(_3\) respectively. The SVM had a MRE of 6.3, 8.3, and 6.6 percent for RH, NO\(_2\), and NH\(_3\) respectively [64]. De Vito et al. compared the performance of a tapped delay neural network and non-linear autoregressive neural model with exogeneous inputs (NARX)\(^5\), which use past and present inputs, to a FFNN, which only uses current inputs. Using TDNN the MRE for NO\(_2\), O\(_3\), and NO\(_x\) was 22, 42, and 20 percent respectively compared to 25, 70, and 29 percent using FFNN. The TDNN model had the lowest MRE, especially for O\(_3\), which implies that there may be some effect from past values on current values [30]. For this study, only electrochemical sensors were used. Other types of RNN should also be explored since the past measurements may affect current predictions.

---

3 TDNN is a neural network which also uses past inputs to predict current outputs
4 SVM is a non-linear model which creates an optimal hyperplane to separate inputs
5 NARX is a neural network which uses past inputs and recursively calculates the
2.4 Evaluation of current sensors

Several commercially available low-cost sensors have been tested as well. On such sensor is the Aeroqual series 500 portable gas sensor for O$_3$ and NO$_2$. The sensor performed well for O$_3$ with a $R^2$ of 0.91 however it had a $R^2$ of only 0.02 for NO$_2$ [14]. Another study evaluated several different low-cost air quality devices at a reference site. For O$_3$, three different devices were tested; AQMesh, CairClip, and Aeroqual SM50. The AQMesh had correlation of 0.39 to 0.45, the CairClip was 0.82 to 0.94, and the Aeroqual was 0.91 to 0.97. For NO$_2$ the CairClip, AQMesh, and Air Quality Egg were tested with correlations of 0.42 to 0.76, 0.14 to 0.32, and -0.25 to -0.22 respectively. The only device that measured NO was the AQMesh which had a correlation of 0.88 to 0.93 when tested. For CO, the AQMesh and Air Quality Egg had correlations of 0.79 to 0.82 and -0.40 to -0.14 respectively. NO$_2$ seems to be very inconsistent among low cost sensors and low cost devices [22]. The Aeroqual O$_3$ and NO$_2$ was used for a field campaign in Montreal, Canada in order to validate land-use regression models [65]. Many low-cost devices are available for consumers however, when tested many do not perform well, especially for measuring NO$_2$. Better testing conditions such as multiple sites and higher concentrations are required to assess the performance of these low-cost devices.

2.5 Conclusions

From the literature, traffic related air pollution can have several adverse health effects and it is important to understand how factors such as proximity to roadways can affect exposure. Reference instruments, although very accurate, are larger and expensive, making them harder to use for some applications. As a result, low-cost gas sensors have been studied as an alternative to using reference instruments. Two popular sensor types used are metal oxide and electrochemical sensors which have different operation principles, advantages, and challenges when applied to pollution measurement. Although these sensors have been explored, it is still not clear on which sensor is better under real applicable conditions. Low cost sensors have many issues such as cross-sensitivity to other gases and meteorological conditions, which is why some studies have combined multiple sensors to measure one pollutant. Several studies mentioned above have shown that calibrating sensors outdoors alongside reference monitors is better than calibration in a laboratory however, these models are generally tested at the same site that they are trained at. Other studies have used non-linear models such as neural networks in order to account for cross-
sensitivity. However, many of these studies have not tested these non-linear models at sites other than the one used for calibration.
Chapter 3

3 Methods

3.1 Device Design

3.1.1 AirSENCE device and sensors

The sensors investigated in this study were housed in a low-cost air pollution measuring device known as AirSENCE which is approximately 10L in volume and 2 kg in weight. The AirSENCE devices were equipped with several metal oxide sensors, temperature and humidity sensors, and one electrochemical sensor. The metal oxide sensors included were all commercially sold for less than $100 US. The sensors selected for inclusion in AirSENCE had previously been optimised over time and sensors that did not show significant results in the earlier versions were omitted from the version of AirSENCE used in this study.

Table 1: List of candidate metal oxide and electrochemical sensors included in AirSENCE devices for this study to measure NOx and O3 concentrations.

<table>
<thead>
<tr>
<th>Sensor Name</th>
<th>Target Gas</th>
<th>Sensing mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHT22</td>
<td>Temperature, humidity</td>
<td></td>
</tr>
<tr>
<td>MQ-3</td>
<td>alcohol, benzene, CH4, hexane, LPG, CO</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>SN706</td>
<td>NOx</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>NO-B4</td>
<td>NO</td>
<td>Electrochemical</td>
</tr>
<tr>
<td>NO2-B4</td>
<td>Temperature</td>
<td>Electrochemical</td>
</tr>
<tr>
<td>MICS-5524</td>
<td>CO, ethanol, hydrogen, ammonia, methane</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>MICS-5526</td>
<td>CO, ethanol, hydrogen, ammonia, methane</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>MICS-4514 Oxidation</td>
<td>NO2</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>MICS-4514 Reduction</td>
<td>CO, hydrocarbons</td>
<td>Metal oxide</td>
</tr>
<tr>
<td>MICS-2614</td>
<td>O3</td>
<td>Metal oxide</td>
</tr>
</tbody>
</table>
The list of sensors used, are shown in Table 1. Signals from the sensors were changes in resistance that correlate to changes in concentrations. These changes in resistance were converted to voltage using a voltage divider circuit. The resistors used in these voltage dividers were selected using the recommended values from the manufacturer datasheets. Voltage outputs from these sensors all had a minimum of 0 and maximum of 5 volts that were digitised to an integer between 0 and 1023. All of the data was averaged to 10 minutes to reduce random noise as well as the influence of residence time in the AirSENCE chamber and any delay resulting from the < 20 m distance between the reference instrument and AirSENCE.

3.2 Device Calibration

3.2.1 Reference instruments

Reference instruments were used in this study in order to calibrate and evaluate the different sensors and calibration models. These “gold standard” instruments were calibrated regularly and believed to represent accurate pollution concentrations. The instruments were: NOx Thermos 42i analyzers for NO and NOx, and Thermos 49c or 49i O3. Environmental variables such as temperature, relative humidity, and pressure as well as traffic count data for local roadways were also measured.

3.2.2 Study sites and devices

Three different sites were used in this study, with different levels of concentration and reference data available for calibration and evaluation. The first site was located at the University of Toronto (43.659, -79.395), at a height of approximately 3 meters, a location with air pollution levels representative of downtown Toronto [66]. NOx levels at this “Downtown Site” generally range from 10 to 20 ppb, but can briefly reach higher values due to heavy emitting vehicles. The typical traffic levels at this site are shown in Table 2. The second site was next to the Highway 401 (43.711, -79.543), North America’s busiest highway, at a height of approximately 10 meters. This site has elevated levels of many traffic related air pollutants including CO, NOx, ultrafine particles, and black carbon. Typical conditions at this site are also shown in Table 2. For
example, NOx at this “Highway Site” often exceeds 80 ppb and at times levels are as high as 250 ppb. Both sites have distributions heavily skewed towards low NOx concentrations. However, due to heavy emitting vehicles, together they experience a wide range of NOx concentrations ranging from below detection to 250 ppb. The third site was Peking University in Beijing, China. This “PKU” site experienced some of the highest concentrations, especially for O3. This site is also located next to a major roadway however the effect from “plumes” of car driving by is less evident than the sites in Toronto. Pollutant concentrations at this site, during the study period, were sustained high for many days, which could imply more of an effect from regional pollution.

Table 2: Summary of conditions at Highway and Downtown sites during the study.

<table>
<thead>
<tr>
<th>Site</th>
<th>5th, 50th, 95th percentile NO concentration (ppb)</th>
<th>5th, 50th, 95th percentile NO2 concentration (ppb)</th>
<th>5th, 50th, 95th percentile O3 concentration (ppb)</th>
<th>Average traffic count (veh/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downtown</td>
<td>0, 2, 19</td>
<td>2, 9, 28</td>
<td>7, 23, 49</td>
<td>17,000</td>
</tr>
<tr>
<td>Highway</td>
<td>0, 11, 86</td>
<td>3, 16, 39</td>
<td>0, 16, 49</td>
<td>400,000</td>
</tr>
<tr>
<td>PKU</td>
<td>0, 8, 117</td>
<td>5, 29, 65</td>
<td>1, 8, 67</td>
<td>No information</td>
</tr>
</tbody>
</table>

Three separate devices were used in this study to determine whether a model calibrated with one device could be applied to several devices at several locations. The first operational device, denoted Device 1, was initially deployed at the Downtown from May 15 to June 6, 2016. From June 6 to July 6, 2016, the device was deployed next to the Highway site. Device 1 then returned to the Downtown Site from July 6 to July 13, 2016 and was deployed to PKU. Reference data from PKU was available from August 15 to August 30, 2016 and again from November 5 to November 30, 2016. The second device, denoted Device 2, was deployed at the Downtown site from October 10 to December 1, 2016. The third device, Device 3, was deployed to the Highway site from July 6, 2016 to February 1, 2017. Device 3 was not equipped with the NO-B4 sensor so it was not used for assessing the performance of the NO models.
3.3 Data Analysis

3.3.1 Cross-responsiveness with other gases and environmental factors

Another major problem with low cost sensors is that they can be cross sensitive to gases other than the gas that they were designed to measure. This is especially apparent for the metal oxide sensor. (Table 3) shows a list of the sensors examined and their cross-sensitivities based on manufacturer specifications [43,67–72]. Alphasense, the makers of the electrochemical NO-B4 sensor used in AirSENCE, tested the cross responsiveness of their sensors to other pollutants. The effects of cross sensitivity due to other pollutants only seem to affect the NO-B4 sensor at concentrations that are generally much higher than ambient conditions so they are in effect negligible [44].

Environmental factors also play a key role in the response of sensors. Temperature, humidity and pressure can have a range of effects on sensor readings. For metal oxide sensors temperature generally increases the baseline and sensitivity of sensor readings. For metal oxide CO sensors humidity can increase the baseline and sensitivity of the sensors. The NO-B4 sensor is very sensitive to temperature, humidity, and pressure. Alphasense uses another electrode which is located away from the sensing electrode to measure the effect of environmental response to the sensor. A previous study [19] shows that when humidity is above 80% the noise of the response becomes higher. When the pressure is not atmospheric, the values of the sensor can dramatically change and take a long time to recover [44].

As a result of the cross-responsivity to both other gases and environmental factors, many sensors are often required to measure one pollutant. There also may be a non-linear response of one sensor to different concentrations of gases and levels of environmental factors, which is why non-linear models were explored in this study.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>gas</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiCS-5521</td>
<td>CO</td>
<td>No information</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>No information</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>VOC</td>
<td>No information</td>
<td></td>
</tr>
<tr>
<td>MQ3</td>
<td>Alcohol gas 0.4mg/l alcohol</td>
<td></td>
</tr>
<tr>
<td>NO2-B4</td>
<td>NO2</td>
<td>No information</td>
</tr>
<tr>
<td>O3</td>
<td>&gt; 500 filter capacity ppm.hr @ 0.5 ppm</td>
<td></td>
</tr>
<tr>
<td>H2S</td>
<td>&lt; -80 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td>&lt;5 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>Cl2</td>
<td>&lt;65 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>SO2</td>
<td>&lt;1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>&lt;3 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>C2H2</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>NH3</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>CO2</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>NO-B4</td>
<td>NO</td>
<td>No information</td>
</tr>
<tr>
<td>H2S</td>
<td>&lt;10 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>NO2</td>
<td>&lt;4 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>CL2</td>
<td>&lt;3 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>SO2</td>
<td>&lt;5 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>&lt;0.3 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>NH3</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>CO2</td>
<td>&lt;0.1 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>O3</td>
<td>&lt;4 sensitivity % measured gas @ 5 ppm</td>
<td></td>
</tr>
<tr>
<td>MiCS-2614</td>
<td>O3</td>
<td>No information</td>
</tr>
<tr>
<td></td>
<td>NO\textsubscript{x}</td>
<td>200 RG/RA</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>SN706</td>
<td>NO\textsubscript{2}</td>
<td>45 RG/RA</td>
</tr>
<tr>
<td></td>
<td>H\textsubscript{2}</td>
<td>-2.2 RG/RA</td>
</tr>
<tr>
<td></td>
<td>NH\textsubscript{3}</td>
<td>-1.1 RG/RA</td>
</tr>
<tr>
<td></td>
<td>CO\textsubscript{2}</td>
<td>No response</td>
</tr>
<tr>
<td></td>
<td>SO\textsubscript{2}</td>
<td>No response</td>
</tr>
<tr>
<td></td>
<td>CH\textsubscript{4}</td>
<td>No response</td>
</tr>
<tr>
<td>MiCS-4514</td>
<td>CO</td>
<td>No information</td>
</tr>
<tr>
<td></td>
<td>NO\textsubscript{2}</td>
<td>No information</td>
</tr>
</tbody>
</table>

3.3.2 Feature selection

Before any feature analysis was done, the sensors were compared to each other to remove correlated sensors since this may affect the feature selection process.
The most correlated sensors were removed so there would be enough variability to assess which remaining sensors should be used in each model. The TGS822, TGS880, MICS5524, and MICS5526 sensors were removed because they had an $R^2$ above 0.8 with another sensor as shown in Figure 3.

To train the models to predict pollution, the best sensors must be selected and the unnecessary sensor must be omitted. There are many different methods for selecting features for inputs in models. It is common to use a stepwise selection process to find the best features for any type of model including neural networks. Forward selection finds the best features by training a model with one sensor, adding another sensor, and keeping that sensor if it increases the metric such as adjusted $R^2$. Another type is backward selection, where a model is made with all features and features are successively removed. There is also stepwise selection, which combines both forward and backward selection. Many studies involving neural networks use these methods for feature selection [73,74]. The method used for this thesis was the Regsubsets function in leaps. This function does an exhaustive search of all features and ranks the combination of features which have the highest adjusted R-squared value [75]. An exhaustive search means that this
function trains thousands of multiple linear regressions using different combinations of features. This type of selection process would take a very long time if neural network models were used for the exhaustive search so therefore it was assumed that using a linear model would give the best features. Another method for ranking feature importance are added-variable or partial regression plot. The point of this plot is to show the effect of removing an independent variable from a model. The y-axis is the residuals from regressing the target pollutant against the all sensors except the sensor of interest. The x-axis is the residuals from regressing the sensor of interest against all other sensors. If there was a relationship in the plot then that sensor was important to the final model.

From regsubsets and the partial regression plots, temperature, MICS-5121, MQ3, and MICS-4514 have a relationship with O$_3$. SN706, NO2-B4, MICS-4514, and MICS-2614 seem to have a relationship with NO$_2$ and there is almost an exclusive relationship between NO and NO-B4. The features from the added-variable plots agree with the highest ranked features from the regsubsets package so these sensors were selected.

Regsubsets and partial regression plots provide a good estimation of which sensors are most important. However, sensors were also added and removed manually, using the previous methods as a starting point. One finding from the feature selection was that the NO-B4 electrochemical sensor was the most important for measuring NO however the NO2-B4 sensor was not as good as the metal oxide sensors for measuring NO$_2$. This means that the electrochemical sensing mechanism is more reactive with NO than NO$_2$ even though many low cost versions mentioned in Chapter 2 have been using the Alphasense NO2-B4 sensor [17,18,29,30].

### 3.3.3 Normalization

Normalization is a very important step when using neural networks. Since the weights of a neural network are randomly initialized, it would take a long time to optimize them if some values had high ranges. Another important reason for normalization is that the sensors between devices may have different baseline and sensitivities, which would make models trained on one device not transferrable to another device. Standardization was used to normalize the values, which is when
the sensor value is subtracted by the mean and divided by the standard deviation. Standardization is the preferred method because if there are some outliers they will not have as large as an effect.

\[ z = \frac{x - \mu}{\sigma} \]  \hspace{1cm} (1)

Equation 1 was the equation used to standardize the sensor values. Z is the standardized value, x is the raw sensor value, \( \mu \) is the mean of the sensor values, and \( \sigma \) is the standard deviation of the sensor values. The first month of sensor values from each device were used for standardization. Ideally, devices would first be collocated before they were standardized to make sure that they experienced the exact same ranges and combinations of pollution and environmental factors however, some devices were deployed at different times than others.
Chapter 4

4 Sensor based measurements of Nitrogen oxides using three array designs

4.1 Calibration and evaluation

Calibration data refers to both raw low cost sensor signals and corresponding reference data, which was used to optimize the weights of a model in order to predict the reference data. Typical lengths of calibration data may vary but two weeks has been found to be enough [21]. Evaluation data is the remainder of the raw sensor signals and corresponding reference data which is used to assess how accurate and precise the model is. Assessing how well a model performs is done using several metrics including R-squared, mean absolute error (MAE), slope, and intercept. The R-squared shows how linear the relationship is between the reference data and the predicted values. MAE is the average absolute difference between the reference data and this is a measure of how accurate the model is. When the reference data and predicted values are regressed, the slope is a measure of the sensitivity of the model and the intercept is a measure of the bias of the model. The best model would have a slope that is close to one and an intercept that is close to zero.

4.2 Metal oxide, electrochemical, and combined sensor array performance at highway site

NOx concentrations at the Highway Site covered a wide range, that often exceeded 150 ppb, as shown in Figure 4. This wide range allowed a comparison of the metal oxide or electrochemical sensor types alone, and whether combining both sensor types into the same calibration model improved predictive performance. All calibration models were trained using the first two weeks of data from this site (June 6 to July 19, 2016) and tested using the last two weeks (June 20 – July 6, 2016).
Figure 4: Box and whisker plot of NOx concentration measured June 20 – July 6, 2016 at the Highway Site. The edges of the box represent the 25th and 75th percentiles, the center line is the median, the whiskers are the maximum and minimum, and the dots are the outliers. The maximum and minimum are defined as 1.5 times the interquartile range above the 3rd quartile and below the 1st quartile and outliers are outside that range. Each of the groups of boxes represents increasing 50 ppb levels of reference NOx data. The electrochemical is closer to the reference measurements than metal oxide sensors, more notably at higher concentrations.

At the lowest levels of NOx, the median values agreed to within 15 ppb but the metal oxide sensors overestimated the maximum of the reference data by up to 25 ppb, excluding the outliers. This overestimation was in part due to cross sensitivity with O3 at low NOx concentrations; below NOx levels of 50 ppb (Figure 5), the metal oxide sensor array data were correlated with O3 ($R^2 = 0.35$). In contrast, the electrochemical sensor was not significantly affected by O3 at higher NOx concentrations. A previous study suggested that the poor performance of NO2 measurements, a component of NOx, using metal oxide sensors was due to cross-sensitivity with O3 and may be offset by combining different sensor types [17].
At 50 to 100 ppb NO\textsubscript{x}, the medians for both sensor types were within 5 ppb of the reference and the RMSE was within 10 ppb (Figure 4). Above 100 ppb, the metal oxide sensor array reached an upper limit and did not linearly follow the reference data. Thus, the electrochemical sensors showed much better agreement with the reference data above 100 ppb NO\textsubscript{x}, and this difference was even more evident above 150 ppb.

Above 100 ppb, the electrochemical sensor was better at measuring NO\textsubscript{x} than an array of metal oxide sensors. However, the metal oxide sensor array showed better agreement between 50 and 100 ppb. Thus, a combination of metal oxide sensors and electrochemical sensors was explored to see if the strength of each sensor type could thereby be leveraged.
The combined array did offer some advantages; the addition of metal oxide sensors did improve measurements below 150 ppb. The median, 1\textsuperscript{st} and 3\textsuperscript{rd} quartiles, maximum, and minimums of the electrochemical sensor and the combined array all, respectively, agreed within 10 ppb for NO\textsubscript{x} concentrations below 150 ppb. Above 150 ppb the combined array under predicted the reference distribution by 20 ppb, due to the metal oxide sensors reaching an upper limit at higher NO\textsubscript{x} levels (Figure 6). Although O\textsubscript{3} affected the metal oxide sensor below 50 ppb of NO\textsubscript{x} (Figure 5) the effect was negligible when electrochemical sensors were also present in the array.

Table 4 summarizes the performance of the three types of sensor arrays investigated. The slope was lowest for metal oxide sensors (Slope = 0.74) since the response was non-linear above 100 ppb. The R\textsuperscript{2} was highest and the RMSE was lowest for the combined sensor array (R\textsuperscript{2} = 0.80,
RMSE = 12.6) since the metal oxide sensors were more sensitive below 100 ppb, which was the level most often measured at the Highway Site. The improved performance provided by including the metal oxide sensors would be even more important at the 10 to 20 ppb NOx concentrations more often encountered across cities in developed western nations. This reasonable performance at these low concentrations provides reason for optimism that sensor based devices will become effective tools for estimating exposure in the cleaner and more sustainable smart cities we aspire for in future.

Table 4: Reference vs. sensor regression statistics for determining the precision and accuracy of sensor based estimates of NOx concentrations up to 250 ppb.

<table>
<thead>
<tr>
<th>Sensor type</th>
<th>Slope</th>
<th>Intercept</th>
<th>R squared</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal oxide</td>
<td>0.74</td>
<td>13</td>
<td>0.68</td>
<td>19.3</td>
</tr>
<tr>
<td>Electrochemical</td>
<td>0.78</td>
<td>11</td>
<td>0.72</td>
<td>13.4</td>
</tr>
<tr>
<td>Combined</td>
<td>0.86</td>
<td>10</td>
<td>0.80</td>
<td>12.6</td>
</tr>
</tbody>
</table>

In summary, the electrochemical sensor was found to be better than the array of metal oxide sensors for predicting NO$_x$ (Figure 4). However, combining the metal oxide sensors and the electrochemical sensor was even more accurate, particularly for measuring low NOx concentrations. At concentrations above 150 ppb the electrochemical sensor alone provided the most accurate data which is due to the ratio of NO to NO$_2$ being higher at higher concentrations of NO$_x$ for ambient data at the Highway site.

4.3 Site selection for sensor training

Sensor-based air quality monitors need to be able to provide reliable data for sites across cities that experience different levels of pollutants. Further, they need to be reliable when calibrated at one site and then deployed to another. This transferability between sites was investigated by comparing the reliability of NO$_x$ measurements based on calibrations at the Highway or
Downtown Sites. Calibrations for each site were developed using two weeks of training data and tested using the rest of the data from both sites combined. The combined array of metal oxide and electrochemical sensors from the previous section were used.

![Graph showing calibration results for Downtown and Highway Sites.](image)

The array calibrated at the Highway Site agreed well with the reference data yielding a slope of 0.9, while calibration at the Downtown Site had a slope of 0.6 (Figure 7). It is not surprising that the Highway Site provided better calibration data given the larger range of NO\textsubscript{x} than the Downtown Site. In general, the sites with a larger range of pollutant concentrations should be selected for device calibration. Intuitively, calibration with data covering the full range of
interest will allow interpolation whereas if the upper or lower range of interest is missing, extrapolation of the calibration will be needed which is more prone to error.

4.4 Training to application site transferability

The main goal of this study was to assess how accurately different types of low cost sensors perform when they are exposed to sites that were not used for training. In order to be confident that sensor-based devices can work under various operating conditions, they must be transferable from the site where they were calibrated to other sites. This transferability was evaluated for the three array options by applying a calibration at the Highway Site to measurements at the Downtown Site. Data from the Highway Site was selected for calibration as it had the widest range and yielded the best calibration (Figure 7). The three array designs were then tested at the Highway and Downtown Sites to compare the relative benefits across multiple sites in terms of the increased costs.

Figure 8: Box and whisker plot for sensors calibrated at the Highway Site and then used for NOx measurements at the Highway or Downtown Sites. The edges of the box represent the 25th and 75th percentiles, the center line is the median, the whiskers are the maximum and minimum, and the dots are the outliers. Spatial difference between the distributions Downtown and at Highway are higher with electrochemical sensors.
As expected, better agreement was obtained when calibrating and then using the sensors at the same site (Figure 8). When trained at the Highway Site and deployed to the Downtown Site, all three sensor array designs yielded positive median residuals indicating over-estimation of the NOx concentrations as compared to the reference data. At the Downtown Site, the metal oxide array had the largest bias, with a median value that was approximately 10 ppb higher than those for the electrochemical or combined sensor arrays. This resulted in the metal oxide sensors over-estimating the “true” NOx concentration by almost a factor of two.

Overall, biases occurred for all three array designs when the devices were deployed to a different site (Figure 6), however the relative error (Figure 9) was lower for electrochemical and combined arrays (RMSE=17.5 ppb, RMSE=17.4 ppb) than a metal oxide sensor array (RMSE=25.2 ppb). When regressed with the reference data, the $R^2$ and slope of the combined array was most ideal ($R^2=0.50$, slope=0.93) compared to the metal oxide array ($R^2=0.46$, slope=1.27) and electrochemical ($R^2=0.27$, slope=0.55). The bias was also lowest for the combined array (bias=12.1 ppb) compared to the metal oxide array (bias=15.1 ppb) and the electrochemical sensor (bias=16.6 ppb). These biases were likely in part due to the different pollutant composition at the two sites. The NOx at the Highway Site was typically 52% NO2 while at the Downtown Site it was 88% NO2. The NO-B4 electrochemical sensor used here had stronger sensitivity to NO than NO2. Thus, it performed surprisingly well when calibrated at a site with higher NO and transferred to a site with low NO. This greater NO sensitivity may also have contributed to its better performance in measuring the higher NOx at the Highway Site. However, even with the addition of metal oxide sensors, the combined array still an overestimated NOx at the Downtown Site. The O3 concentrations were lower beside the highway due to titration by the excess NO present. Thus, this potential interference was also presumably also higher at the Downtown Site, which may have inflated the values measured by the metal oxide array.
Figure 9: Performance trade-off of the metal oxide sensors, electrochemical and combined array designs in the measurement of NO\textsubscript{x}. Relative error is calculated as the average of the reference values minus the predicted values divided by the reference values.

Although the cost of the metal oxide sensor array was ~$60 US less than the electrochemical sensors, they exhibited a 7 ppb higher RMSE at both the training site and the testing site. This is significant relative to the 10 to 20 ppb NO\textsubscript{x} concentrations often measured at the Downtown Site. Relative error values were calculated so as to illustrate these relative magnitudes. In particular, the metal oxide sensors alone exhibited very large relative errors so it appears that the extra cost of including the electrochemical sensor is worth the RMSE reduction. The NO-B4 sensor used in this study is one of the most popular electrochemical sensors on the market. Using a metal oxide sensor array does offer advantages for measuring pollutants other than NO. The most expensive component of the NO electrochemical sensor is the individual sensing board, which is required to filter and amplify the current obtained from the sensor. If this could be directly built into the device circuit board, it would reduce the cost of the electrochemical sensor to almost the same price as the metal oxide sensor array. However, sensors within this metal oxide array can also be combined with other sensors to measure other pollutants.
Chapter 5

5 Comparison of different models used for calibration of low-cost sensors

In this chapter, non-linear models are compared to linear models for calibrating low-cost sensor arrays. These models were developed using one device trained at one site and tested across many devices and sites to assess how well they perform outside of the training conditions.

5.1 Deep neural network

A single layer feed forward neural net (FFNN) is the simplest type of neural network, where input values are passed through a hidden layer and then output as one or more predictions. The hidden layers consist of n hidden nodes and have an activation function which can be linear, sigmoid, tan, etc. Before inputs are passed through hidden nodes they are multiplied by weights which are randomly generated and optimized using backward propagation. Backward propagation is a process by which outputs are compared to the “real” target values and then the estimated optimizations to the weights are passed backward through the network. Unlike a linear regression, a neural network can handle non-linear problems as well.

A deep neural network (DNN) is like a FFNN except that it consists of more than one hidden layer. Outputs from one hidden layer are passed as inputs into another hidden layer and optimized using backward propagation. Deeper hidden layers are ideally used to refine the features of the input data. One of the biggest issues with the deep hidden layers is the vanishing or exploding gradient [76]. Exploding gradients mean that gradients, or amount of change, is larger in the first layers, which means that only the first layers are optimized and the weights of the later layers are almost unaffected. The opposite and more common problem is the vanishing gradient. This is where the last layers in the DNN are optimized faster than the first layers by backward propagation. If these problems occur then there is no point with having multiple hidden layers since only one layer will be optimized anyway. For this thesis, the program used to train the models was the python-based TensorFlow system [77]. In order to mitigate the effect of vanishing gradient a ReLU activation function was used [78,79]. The optimizer used was AdamOptimizer [80] with a learning rate of 0.001 and the cost function was reducing the mean of the square of the weights. This reduces the cost or error of the model by adjusting the weights.
Next, the optimal number of layers and nodes for each layer was estimated. As a rule of thumb, the number of nodes in each layer is generally around the number of input features; however, adding more nodes typically does not negatively affect the model but may increase computational time [81]. The number of nodes in each layer was kept the same since, in some studies, this works better [81]. In order to find the best number of nodes, a list of 5, 10, 20, 50, 100, 150, 200, 250, and 300 nodes were tried for each layer. Initially, as the number of nodes were increased, the mean squared error (MSE) decreased, however, after approximately 50 nodes in all layers, the MSE began to increase. Therefore, the final number of nodes in each hidden layer was 50. All additional parameters for the neural network are shown in Table 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden nodes in each layer</td>
<td>50</td>
</tr>
<tr>
<td>Hidden layers</td>
<td>2</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Cost function</td>
<td>Equation 1</td>
</tr>
<tr>
<td>Optimizer function</td>
<td>AdamOptimizer</td>
</tr>
</tbody>
</table>

5.2 Recurrent neural network and Long Short-Term Memory

A recurrent neural network (RNN) is similar to a FFNN except that the hidden state from the previous time step is included with the current inputs. This also means that, as the model is being trained, information from the past cascades through time to the present. As opposed to FFNN, which uses back-propagation, RNN uses backpropagation through time. This is similar to backward propagation, however they have to link back to the previous hidden state. Just like the
DNN, RNN can also have problems due to vanishing and exploding gradients. However, rather than the gradients affecting the layers unevenly they affect the different times differently. With an exploding gradient, outputs are more affected by inputs from several time steps before. With a vanishing gradient, only the current or recent inputs affect the output and events from the past will not have any affect.

In order to reduce the effect from exploding and vanishing gradient, the long short-term memory (LSTM) cell was used [82–84]. The LSTM uses different gates to choose which past inputs are to be used for the current measurement. These cells consist of three types of gates, an input gate, an output gate, and a forget gate. Each gate receives the inputs of the current time and the output from the previous time and they have a matrix of weights which helps decide whether the gate should be open or not. Also, instead of the multiplying the change in weight by the weight, the change in weight is added to the weight which reduces the effect of exploding and vanishing gradient through the time.

The RNN used for this study also uses the LSTM cell, which accounts for the vanishing gradient problem. RNN was selected since there may be non-linearity in the data and there may be some time dependence or a delayed response between the sensors and the current measurements. Combining the ability of the model to leverage non-linearity and time dependence across an array of sensors may improve measurements and be more accurate when applied to different sites and seasons. The RNN was also calibrated using the package TensorFlow. This type of model uses the same cost and optimizer function as the deep neural network. The main difference between them is that the recurrent neural network uses a LSTM cell rather than a single node and there are not multiple layers like in a deep neural network.

5.3 Comparison of MLR, DNN, and RNN for measuring air pollution using low-cost sensor array

The data used in this chapter was described in Section 3.2.2 and a summary is provided in Table 6. Device 1 has been separated into Toronto and PKU in order to compare the differences of measuring air pollution in different cities. One major challenge with low-cost sensors is that different sensors may have different resistances in “zero” air and may have different sensitivities
to increasing gas concentrations or temperatures. Ideally, devices would be collocated in order to
generate the normalization coefficients described in Section 3.3.3. However, for this study
devices were not collocated, therefore, one month of data was used from each device in order to
generate the normalization coefficients. The idea was that over one month the distribution of the
reference or “actual” pollution concentrations would be close enough between devices. Then, a
mean and standard deviation of the sensors would represent the differences between devices, not
environments. The distribution of O$_3$ and NO$_2$ is quite similar between the Downtown and
Highway sites and since Device 3 could not measure NO, the difference in NO distributions did
not matter.

<table>
<thead>
<tr>
<th>Device</th>
<th>Location</th>
<th>Start date</th>
<th>End date</th>
<th>Pollutants measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device 1</td>
<td>Downtown, Toronto</td>
<td>May 13, 2016</td>
<td>June 6, 2016</td>
<td>NO, NO$_2$, O$_3$</td>
</tr>
<tr>
<td></td>
<td>Highway, Toronto</td>
<td>June 6, 2016</td>
<td>July 6, 2016</td>
<td>NO, NO$_2$, O$_3$</td>
</tr>
<tr>
<td>Device 1</td>
<td>Peking University, Beijing</td>
<td>August 15, 2016</td>
<td>November 30, 2016</td>
<td>NO, NO$_2$, O$_3$</td>
</tr>
<tr>
<td>Device 2</td>
<td>Downtown</td>
<td>October 10, 2016</td>
<td>December 1, 2016</td>
<td>NO, NO$_2$, O$_3$</td>
</tr>
<tr>
<td>Device 3</td>
<td>Highway</td>
<td>July 6, 2016</td>
<td>February 1, 2017</td>
<td>NO$_2$, O$_3$</td>
</tr>
</tbody>
</table>

Calibration models were trained with the first two weeks of data from Device 1 (Toronto) when
it was located at the Downtown site. The same models calibrated with Device 1 were applied to
Device 2 and Device 3 in order for the evaluation to be as rigorous as possible. From Figure 10,
the median of the DNN model was the closest to the 1:1 line for all pollutants, however, the
range of the 10$^{th}$ and 90$^{th}$ percentile were larger than for MLR and RNN. For measuring NO with
Device 1 (Toronto), the $R^2$ was above 0.84 for all models but the DNN had the best slope, MAE, and $R^2$ (slope=0.75, MAE=6.2, $R^2=0.88$) followed by the RNN and MLR models (Table 7). The DNN model could measure NO concentrations up to 2 times higher than observed in the training dataset.

When the models obtained by calibrating Device 1 at the Downtown site were applied to Device 2, the results were just as good as for Device 1 (Toronto), with the DNN having a slope of 0.96 compared to 0.81 for the RNN and 0.52 for the MLR (Table 7). However, all models had a much lower slope when applied to Device 1 (PKU) data since the NO concentrations that were consistently higher than the concentrations during the testing period.

When measuring NO$_2$, the MLR model had the highest slope for Device 2 (slope=0.77) and Device 3 (slope=0.62) but for Device 1 (Toronto), the DNN had the highest slope (0.75). When applied to Device 1 (PKU), none of the models worked well and the predictions were not correlated with the reference data ($R^2<0.1$) (Table 8). From Figure 10, the median of the DNN was closest to the 1:1 line however, the 10$^{th}$ to 90$^{th}$ percentile had a much higher range than other models. For measuring NO$_2$, it is hard to tell which is the best model using the statistics from Table 8 however, from Figure 10, the DNN model appears to at least be able to measure higher concentrations.

For O$_3$, all models had a similar slope, $R^2$, and MAE when applied to Device 1 (Toronto) and Device 2. When applied to data from PKU, the slope using DNN was 0.76 compared to 0.65 using MLR and 0.57 using RNN. From Figure 10, for O$_3$, predictions from all models begin to deviate lower from the 1:1 line at different concentrations. The DNN began to deviate from the 1:1 line at ~80 ppb and the RNN and MLR deviate at ~60 ppb. Although the DNN model had the highest slope, 0.96, the RNN model had the highest $R^2$ of 0.71 (Table 9).
Table 7: NO performance against reference data for all devices, excluding Device 3, using calibration data from Device 1 (Toronto) at Downtown and Highway sites.

<table>
<thead>
<tr>
<th>Device</th>
<th>Calibration site</th>
<th>Model</th>
<th>R²</th>
<th>MAE</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device 1</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.85</td>
<td>11.1</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.88</td>
<td>6.2</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.86</td>
<td>8.0</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.79</td>
<td>7.3</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.84</td>
<td>5.7</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.82</td>
<td>5.8</td>
<td>0.87</td>
</tr>
<tr>
<td>Device 1</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.99</td>
<td>27.0</td>
<td>0.29</td>
</tr>
<tr>
<td>(PKU)</td>
<td></td>
<td>DNN</td>
<td>0.99</td>
<td>16.0</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.96</td>
<td>20.2</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.99</td>
<td>11.1</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.99</td>
<td>8.6</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.97</td>
<td>9.3</td>
<td>0.69</td>
</tr>
<tr>
<td>Device 2</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.73</td>
<td>2.5</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.74</td>
<td>2.3</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.74</td>
<td>2.2</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.72</td>
<td>7.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.76</td>
<td>5.0</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.75</td>
<td>5.2</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Table 8: NO₂ performance against reference data for all devices using calibration data from Device 1 (Toronto) at Downtown and Highway sites.

<table>
<thead>
<tr>
<th>Device</th>
<th>Calibration site</th>
<th>Model</th>
<th>R²</th>
<th>MAE</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device 1 (Toronto)</td>
<td>Downtown</td>
<td>MLR</td>
<td>5.8</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>6.1</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>6.7</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td>MLR 0.52</td>
<td>5.3</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNN 0.48</td>
<td>5.4</td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNN 0.57</td>
<td>4.7</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device 1 (PKU)</td>
<td>Downtown</td>
<td>MLR</td>
<td>36.2</td>
<td>-15.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>23.2</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>27.5</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td>MLR 0.52</td>
<td>18.7</td>
<td>-5.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNN 0.14</td>
<td>16.7</td>
<td>32.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNN 0.48</td>
<td>14.7</td>
<td>4.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device 2</td>
<td>Downtown</td>
<td>MLR</td>
<td>5.6</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>9.1</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>8.9</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td>MLR 0.58</td>
<td>5.5</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNN 0.52</td>
<td>8.5</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNN 0.65</td>
<td>4.9</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device 3</td>
<td>Downtown</td>
<td>MLR</td>
<td>8.0</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>7.2</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>6.8</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td>MLR 0.44</td>
<td>12.4</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNN 0.49</td>
<td>6.3</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNN 0.47</td>
<td>9.8</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9: O$_3$ performance against reference data for all devices using calibration data from Device 1 (Toronto) at Downtown and Highway sites.

<table>
<thead>
<tr>
<th>Device</th>
<th>Calibration site</th>
<th>Model</th>
<th>$R^2$</th>
<th>MAE</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device 1 (Toronto)</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.9</td>
<td>4.5</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.89</td>
<td>5.2</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.88</td>
<td>4.7</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.84</td>
<td>4.9</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.85</td>
<td>4.5</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.80</td>
<td>5.9</td>
<td>0.72</td>
</tr>
<tr>
<td>Device 1 (PKU)</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.87</td>
<td>9.3</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.91</td>
<td>8.6</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.82</td>
<td>11.2</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.89</td>
<td>9.8</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.86</td>
<td>14.4</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.84</td>
<td>12.7</td>
<td>0.41</td>
</tr>
<tr>
<td>Device 2</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.85</td>
<td>10.4</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.88</td>
<td>9.8</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.84</td>
<td>9.5</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.82</td>
<td>6.1</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.77</td>
<td>6.3</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.81</td>
<td>4.6</td>
<td>1.15</td>
</tr>
<tr>
<td>Device 3</td>
<td>Downtown</td>
<td>MLR</td>
<td>0.66</td>
<td>13.8</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNN</td>
<td>0.58</td>
<td>17.3</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNN</td>
<td>0.71</td>
<td>12.0</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Highway</td>
<td>MLR</td>
<td>0.75</td>
<td>7.6</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>DNN</td>
<td>MLR</td>
<td>RNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>16.6</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNN</td>
<td>0.82</td>
<td>6.0</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10: Comparison of DNN, MLR, and RNN for testing data from all devices calibrated using two weeks of data from Device 1 (Toronto) at the Downtown Site. The line represents the median and the shaded area represents the 10th and 90th percentile of the data.

5.4 Comparison of different training sites for calibrating low-cost sensor array

In order to assess how much of a difference calibration with different sites has, Highway data was used since it generally has higher concentrations of traffic related pollution. Similar to the previous section, two weeks of training data was used from Device 1 (Toronto) at the Highway and applied to Device 2 and Device 3. In Figure 11, the models have very similar medians, 10th and 90th percentiles for NO. In contrast to the models trained Downtown, the medians for NO\textsubscript{2} and O\textsubscript{3} were slightly lower when trained at the Highway. For NO the MLR, DNN, and RNN had slopes that were between 0.87 and 0.92 and R\textsuperscript{2} that were between 0.79 and 0.84 for Device 1 (Toronto). For Device 2, the MLR, DNN, and RNN over predicted values at the Downtown Site with slopes between 1.2 and 1.25. However, for Device 1 (PKU), the calibration models under
predicted concentrations, with slopes between 0.36 and 0.45, much less than models trained Downtown. The DNN had a slope of 0.75 when calibrated using Highway data compared to a slope of 0.62 when calibrated using Downtown (Table 7).

When NO$_2$ was calibrated with Highway data, the slopes were between 0.17 and 0.26 lower than when calibrated with Downtown data for Device 1 (Toronto) but within 0.07 for Device 2 and Device 3. The difference between the R$^2$ of the devices was negligible except that the RNN and MLR performed much better at PKU with R$^2$ that were 0.48 and 0.46 higher and MAE that were 12.8 and 17.5 ppb higher respectively (Table 8). From Figure 11, the median of the DNN was not as close to the 1:1 line as in Figure 10 however, the range between the 10$^{th}$ and 90$^{th}$ percentile was less.

When O$_3$ was calibrated using Highway data, all of the models had slopes that were lower for all sites and models. The slope for Device 2 was between 1.1 and 1.3 when calibrated with Highway data compared to 1.5 for all models when calibrated using Downtown data. The MAE was approximately 6 ppb lower for device 3 for the MLR and RNN models but there was not much change for the DNN (Table 9). From Figure 11, the point where the median begins to deviate from the 1:1 line is ~40 ppb compared to ~80 ppb for the DNN, which was trained using Downtown data.
5.5 Effect of normalizing reference data for training models

In order to help the DNN and RNN converge faster, pollutant concentrations from the reference instruments (targets) were normalized using data from the first month of Device 1 (Toronto) deployment at the Downtown site. For this method, the data used to normalize the reference concentrations should be from the site used to calibrate the devices. The goal was to have normalized concentrations which are close to 0 when they are used for calibration but they do not have to be between 0 and 1 to improve the model performance. Then these coefficients were applied to all reference data and used to unnormalize the AirSENCE calibration model values. For calibrating the MLR, normalizing the target did not have any advantage since the weights of the MLR are calculated not optimized for. However, when calibrating the DNN and RNN there was a noticeable decrease in time that it took to run the model. DNN predictions when calibrated with normalized reference pollutant concentrations did not vary too much from unnormalized model predictions. Weights were randomly initialized between 0 and 1 so by normalizing the reference data, all input features and target features were close to the same scale as the weights. The first two weeks of data from Device 1 when it was at the Downtown device was used to calibrate all the devices.
From Figure 12, the median of the predicted pollutants with the RNN model was higher when the targets were normalized. When measuring NO, the RNN trained at Downtown had a slope of 1.0 for Device 1 (Toronto) and 0.9 at PKU compared to 0.6 and 0.45 when trained using unnormalized targets. For NO$_2$, the RNN model had a higher slope, R$^2$, and lower MAE except for Device 3 when using normalized pollutants. From Figure 12, the median was much higher for normalized data but the range between the 10$^{th}$ and 90$^{th}$ percentiles was almost twice as large. However, the 10$^{th}$ percentile of the normalized model is still higher than the 10$^{th}$ percentile of the unnormalized model. For O$_3$, the RNN model performed much better for Device 1 (PKU) with a slope of 0.86 compared to 0.57 with unnormalized pollutants. For device 3, the slope also improved from 0.8 with unnormalized pollutants to 1.0. In Figure 12, the median of the O$_3$ with normalized pollution data was above the 1:1 line until approximately 100 ppb rather than 50 ppb. Although there is a larger range between the 10$^{th}$ and 90$^{th}$ percentile, the RNN trained with normalized pollution could measure much higher concentrations of O$_3$ than even the DNN which was trained using Downtown data.

![Figure 12: Comparison of RNN with normalized target variables and unnormalized target variables.](image)

Figure 13 shows the RNN models trained using two weeks of data from Device 1 (Toronto) at the Downtown site with normalized and unnormalized target values at each site. The model
trained with normalized targets predicts higher values for each device and each pollutant. The worst performing device was Device 1 (PKU) because some of the metal oxide sensors used experienced interference from the environment in Beijing. From Figure 12, RNN trained with normalized pollution data is seems to be able to measure higher concentrations than either MLR or DNN across all pollutants. The training site also seems to have a smaller effect on the model when using a RNN with normalized pollution data. From Figure 13, the NO$_2$ was under predicted at the PKU site for both models.

![Figure 13: Comparison of RNN with normalized and unnormalized target at all sites trained using two weeks of data from Device 1 (Toronto) at the Downtown site. Device 1 (Toronto) was tested at both the Downtown and Highway site, Device (PKU) was tested at Peking University, Device 2 was tested at the Downtown site, and Device 3 was tested at the Highway site.](image)

5.6 Model performance with cross-sensitivity

When using the NO-B4 sensor to measure NO, the cross-sensitivity with other pollutants is very low however temperature effects the model. The auxiliary electrode of the sensor measures the effect of temperature on the model. Therefore the non-linear models only account for cross-sensitivity to temperature. When trained at the Downtown site, the RNN can extrapolate data and predict concentrations up to four times higher than the MLR model can. This is only when the RNN is trained using reference pollution data that are normalized.
For measuring NO\textsubscript{2} achieving high accuracy across multiple devices and sites seems unlikely with the tested sensors. However, it is possible to measure higher concentrations with an RNN model than a DNN or MLR model. Through the feature selection process it turned out that an array of metal oxide sensors was better than electrochemical sensor for measuring NO\textsubscript{2}. NO\textsubscript{2} sensors are very cross sensitive to O\textsubscript{3} which is why the MICS-2614 sensor was used to correct the model. The addition of a temperature sensor did not improve the models in Toronto very much although many of the metal oxide sensors seemed to be affected by the long-term baseline temperature during the Device 3 deployment. Since many sensors used in the model were sensitive to temperature it is possible that they corrected each other for temperature.

For measuring O\textsubscript{3} the MICS-2614 was the best sensor. The main problem with this sensor is that it is inversely proportional to O\textsubscript{3}, so when concentrations are high, the sensor values approach 0. As a result, the values were transformed using a natural logarithm. This sensor did decrease in baseline and sensitivity as the temperature became colder during the device 3 deployment. The sensor also seemed to be correlated with the MQ-3 sensor which measures alcohols and VOCs but it was not correlated with NO\textsubscript{2}. It appears the sensor was also drifting however, when time was added as a variable, the model was less accurate than without time. Normal linear models were only able to measure concentrations of O\textsubscript{3} as high as 60 ppb across all sites however the RNN model could measure concentrations up to 160 ppb which are almost twice as high as concentrations from the training site.

5.7 Model performance at different training sites

From Figure 14, the distribution of NO at the Downtown site was a lot more skewed than the Highway site, which is why the models were not able to measure higher concentrations when trained there. The non-linear models were likely able to measure higher concentrations because the cost function that was used reduced the square of the error. Therefore, the higher points had a higher effect on the cost optimization.
Many points at the Downtown site were under predicted which may be due to the fact that some high pollution events are so short lived that the sensor does not have time to respond. Generally, as individual cars drive by and emit NO plumes, the reference instrument measures these plumes instantaneously. However, the exchange of air in the device is not instantaneous and therefore a plume will be mixed with non-plume air.

The models for calculating O$_3$ generally worked better when trained at the Downtown site. From Figure 15, the distribution of O$_3$ Downtown is less skewed than the distribution at the Highway. O$_3$ gets much lower at the Highway site because NO gets much higher and reacts with O$_3$. As a result, there are more training points at lower concentrations so the model coefficients are also skewed to measure lower concentrations. Also, values measured using DNN and RNN trained Downtown have a lower limit of $\sim$10 ppb since the concentrations of O$_3$ rarely go below 10 ppb at that site.
Figure 15: Distribution of reference $O_3$ for calibration data Downtown and at the Highway compared to the distribution of all data from Device 1.

From Figure 16, the distribution of NO$_2$ at the Downtown site is slightly more skewed than the Downtown site. The models trained with data from Downtown site have higher slopes but lower $R^2$ values. NO$_2$ measurements were the worst for any model and any training site. The highest achievable $R^2$ was 0.65 with Device 2.
Figure 16: Distribution of reference $\text{NO}_2$ for calibration data Downtown and at the Highway compared to the distribution of all data from Device 1.
Chapter 6

6 Applications of low-cost air quality devices

6.1 Highway gradient study

Health effects may vary greatly based on proximity to roadways. Therefore, one useful application of low-cost air quality devices is to accurately measure the difference in pollution at various distances away from the roadway. In February 2017, AirSENCE was collocated alongside reference gas monitors in a campaign to measure the difference between traffic related pollution beside the Highway 401 and 200 meters away from the Highway 401. The monitoring station next to the highway was the same site as described in Chapter 3 and the reference instruments 200 meters away from the highway were the same type. The pollutants measured were NO, NO₂, and O₃ at 5-minute resolution. The calibration models used to measure the concentration were the same models developed in the previous chapter but the RNN were focused on since they performed the best. The devices used for this study used the same models that were calibrated for Device 1 using the first two weeks of data from the Downtown site. Therefore, no data from this study was used to influence the model in any way. The values were normalized using one month of data from when the devices were initially deployed at the Downtown site.

6.1.1 NO results

Figure 17 shows an hourly timeseries of the NO concentrations obtained from the two AirSENCE devices located next to the Highway 401 and 200 meters away from the highway. The device located next to the highway had a much higher concentration than the device 200 meters away so this device is indeed able to measure some spatial difference. When the values between the devices are similar, that is generally because the wind was blowing in the opposite direction from the highway, which is consistent with low pollution times at both sites.
In Figure 18, the AirSENCE measured values were compared to the reference NO values to evaluate the device accuracy. The AirSENCE values are generally slightly lower than the reference during this period, especially the device located 200 meters from the highway which had a median which approximately 5 ppb lower than the reference. Despite this, the AirSENCE values perform quite well with the RNN that was trained at the Downtown site which experienced much lower concentrations. The concentrations from the RNN were output in normalized form and needed to be unnormalized using the coefficients determined in Section 5.6. The model used was not only able to measure higher concentrations but also account for the much colder temperatures experienced during this campaign.
In Figure 19, the difference between NO beside the highway and 200 meters away from the highway were compared based on the reference and AirSENCE data. Most of the points are positive, indicating that the site next to the highway was usually higher. The differences between the reference and AirSENCE were linearly correlated ($R^2 = 0.86$) however the slope was approximately 1.2 therefore the AirSENCE was over predicting the difference. However, it was clear that AirSENCE could reliably measure to within 20%, this spatial gradient due to traffic related pollution.
6.1.2 NO₂ results

NO₂ is another important traffic related pollutant but does not vary spatially as much as NO does. The device located next to the highway was able to measure the concentrations relatively well ($R^2=0.55$, MAE=9.3). From Figure 20, the device located away from the highway did not perform that well. The $R^2$ was high and the slope was good ($R^2=0.70$, slope=0.84), however the intercept was 23 ppb for the device 200 meters away from the highway. The intercept was high because one of the sensors (MICS-4514) used in the model had much higher values than the conditions under which it was trained. This sensor was used to measure NO₂ however, even though it is not mentioned in Table 3, it is also cross responsive to O₃. The values were in the “acceptable” range of concentrations for NO₂ however, the baseline values seem too high. In order to correct these values prior knowledge about the monitoring site would be required.

![Figure 19: Linear relationship in the difference in NO concentrations next to highway and 200 meters away from highway with line of best fit and confidence intervals.](image-url)
From Figure 21, the device located away from the highway had some bias, however, the interquartile ranges were almost the same as the reference. The device located next to the highway had a much larger interquartile range than the reference and the median was approximately 10 ppb lower. Close to the roadways, the AirSENCE had lower $R^2$ but further away from the roadway the $R^2$ increases. This could be due to some interference from other gases produced by vehicle emissions which the sensor surface is cross-sensitive to.
In Figure 22, NO and NO\textsubscript{2} were added together to show total NO\textsubscript{x} concentration at the sites beside the highway and 200 meters away from the highway. The device located away from the highway was about 10 ppb higher than the median of the reference however, the difference is not as high as NO\textsubscript{2} alone. By measuring total NO\textsubscript{x} instead, the issue of poor NO\textsubscript{2} can be mitigated slightly. Although the interquartile ranges of the AirSENCE were not very comparable to the reference, the peaks were within 10 ppb.
6.1.3 $O_3$ results

From Figure 23, peak $O_3$ concentrations for the device located next to the highway were higher than the device located away from the highway and some values were lower than 0. From Figure 24, the reference monitors show that $O_3$ was lower beside the highway, presumably since it gets titrated by NO. The RNN showed that the median $O_3$ concentration was lower next to the highway than 200 meters away and performed better than the MLR that showed the median was higher next to the highway. The device located away from the highway had a median which was approximately 10 ppb lower than the reference data however the interquartile range was similar. Although the performance of the RNN next to the highway was not very consistent with the reference data, it did outperform the MLR model. If average values were taken from the models trained using the RNN, the data would be more useful in showing spatial differences than the MLR model.
Figure 23: Timeseries of AirSENCE O3 measurements next to Highway 401 and 200 meters away from Highway 401.
6.2 Train study

Another application of pollution monitors is to measure pollution levels indoors. Jeong et al. conducted a study where reference instruments were used to measure concentrations of ultrafine particles (UFP) and black carbon (BC) over several months in commuter trains in Toronto, Canada [85]. During one of these trips a NO\textsubscript{x} reference analyzer and AirSENCE device were also included on the train. The following sections show how well the AirSENCE device performed for measuring NO and NO\textsubscript{2} in an indoor environment with elevated concentrations. The device used for this study used the same models that were calibrated for Device 1 with data from the Downtown site only.
6.2.1 NO results

From Figure 25, the hourly concentrations on the commuter train reach levels around 500 ppb which is approximately twice as high as values measured at any of the evaluation sites used in Chapter 5. The environmental conditions in the train were very stable, temperature measured using AirSENCE was around 23 °C +/- 1 °C and the humidity did not fluctuate either. The sensor used to measure O₃ also did not fluctuate very much indicating that O₃ was low, however there was no reference data for this pollutant. The RNN model derived from calibration using the much lower concentrations at the Downtown site performed well when predicting the baseline concentration up to around 150 ppb, however, it did not yield concentrations higher than 200 ppb. By contrast a linear calibration which was developed using “standard” gas in the laboratory performed quite well. This was trained by blowing a mixture of NO and zero air on the sensor surface at concentrations of 0 ppb, 100 ppb, 200 ppb, 400 ppb, and 800 ppb. The sensor voltage levels were then regressed against the known concentrations of “standard” gas to get the sensitivity. The air inside the train does not have the same pollutant mixtures that are observed at the Downtown or Highway sites and there is less fluctuation in temperature. As a result, the laboratory calibration is likely better when applied to concentrations well above those in training data.
6.2.2 NO$_2$ results

From Figure 26, the RNN model was inversely proportional to the actual reference concentration. This is because a key sensor (SN706) used to measure NO$_2$ is highly correlated with O$_3$. Clearly, calibration models developed under ambient conditions next to roadways can yield very erroneous results when applied under other conditions.
In Figure 27, the RNN model for NO₂ was trained without the SN706 sensor, which produced much higher readings during the train study than under ambient conditions. Without that particular sensor, the values were almost flat at 15 ppb for the entire trip even though NO₂ reached hourly concentrations 40 ppb. The value of 15 ppb did match the reference instrument when the NO₂ concentrations in the train would have been due to ambient air as the train was in the push mode (e.g. between 3:0 and 5:00 am).
The goal of low cost air pollution devices is to be able to easily and accurately measure air pollution in a wide range of settings. Some of these applications could be urban air quality such as next to roadways and other applications could be indoor areas where many people spend time like in a train or a parking garage. Many low-cost sensors and devices are not tested rigorously enough to be confident that they will work when deployed in a completely new environment.

The AirSENCE was deployed alongside reference instruments for two major pollution studies including at various distances from a major highway and in a commuter train. For the highway gradient study, the AirSENCE device, using a RNN model, was able to measure the difference between being right next to the highway and 200 meters away from a highway. For NO\(_2\) the device located away from the highway experienced a bias which lead to overestimations of concentration. For O\(_3\), the device located next to the highway had a much higher interquartile range than the reference instrument. However, the difference in median of O\(_3\) between the two devices was like the reference. When the device was put on a commuter train, the NO model performed well up until about 150 ppb even though actual concentrations reached above 500 ppb. The model which used calibration gas was better. NO\(_2\) measured by the device was inversely proportional to the reference data because the model depends highly on O\(_3\).
The NO results are likely better on average because the sensor used (NO-B4) is a very accurate sensor which has been supported by the literature. In ambient settings, the RNN was able to extrapolate data because it accounted for the interactions between the sensor and temperature at different times. During the train study, one of the sensors used to measure NO$_2$, SN706, had an average value that was approximately four times higher than under ambient conditions, which influenced the model. The sensor is responsive to O$_3$, which has an opposite effect as NO$_x$, therefore, since O$_3$ may have been much lower in the train, the baseline resistance of the sensor shifted upwards. By removing the sensor, the model was within the range of “normal” concentrations however it remained relatively flat since the remaining sensor was supposed to measure O$_3$. Under ambient conditions, sensor response to O$_3$ and NO$_2$ are inversely proportional, but since O$_3$ was likely low during the train study, there was no response from the sensor. NO$_2$ is very difficult to measure under ambient conditions whether using an electrochemical sensor or metal oxide sensors. Either sensor type experiences high cross-sensitivity to O$_3$, which the models over fit for when trained under ambient conditions. As a result, when O$_3$, which usually has an inverse relationship with NO$_2$, does not fluctuate the models do not work. Therefore, when measuring NO$_2$ it is important to know what environment is to be investigated.
Chapter 7

7 Conclusion and future directions

7.1 Conclusion

When selecting low cost sensors for measuring NO\textsubscript{x}, there are many factors to be considered including accuracy, precision, price, durability, and lifetime. In this study, electrochemical and metal oxide sensors were compared at sites of contrasting nature to see how well three different sensor arrays designs worked. In order to test whether one model could be applied to multiple devices, models were trained with only one device and applied to all devices in the network. When these arrays were calibrated and tested at a Highway Site, the electrochemical sensor had higher accuracy than the metal oxide sensor array for measuring elevated NO\textsubscript{x} concentrations above 100 ppb. It was found that the metal oxide sensor array plateaued at an upper limit of 120 ppb using a linear calibration model. Even though the metal oxide sensor array was less expensive, the electrochemical sensor studied was more effective but showed deficiencies when measuring low NO\textsubscript{x} concentrations below 50 ppb. A combined array of the electrochemical sensor and several metal oxide sensors was therefore examined. This combined array did show somewhat better performance (higher R\textsuperscript{2} and lower RMSE) for concentrations below 150 ppb. Above this level, the combined array under predicted NO\textsubscript{x} concentration while the electrochemical sensor alone did not. The combined array was approximately twice as expensive as the standalone electrochemical sensor and only slightly more accurate. Thus, if only elevated concentrations above 150 ppb are of interest then the extra metal oxide sensors are not necessary.

In order to be confident that low cost sensors will work at many locations, they must be tested at different sites than where they are calibrated. Ideally, training data should be selected using the largest range of NO\textsubscript{x} concentrations. In this study, arrays calibrated at a Downtown Site with lower NO\textsubscript{x} concentrations under predicted values at the Highway Site with higher concentrations. Arrays trained at the Highway site over predicted concentrations below the third quartile (25 ppb) at the Downtown site but did not over predict the peaks. The electrochemical sensor and combined sensor showed better agreement across the sites than the array of metal oxide sensors alone. Overall, when selecting sensors for a low-cost air quality monitoring device,
the selection of sensors included should be tailored to the desired application and testing needs to
done at multiple sites covering the related range of conditions.

Compared to the MLR and DNN based calibration, the RNN was more effective for calibrating
an array of sensors. Although the RNN did not always have the highest $R^2$ or lowest MAE, it did
have the highest median, 10th and 90th percentiles which means that on average it will predict the
higher episodes of pollution which is more important for health implications. However, before
training the RNN model the target variables should be normalized. Training at different sites did
have an effect on the models, especially for NO. The MLR and DNN models under predicted
concentrations when trained at the Downtown site but less so when they were trained at the
Highway site which had higher average and peak concentrations. This effect of training at
different sites could be mitigated by using the RNN calibration model, which was able to predict
concentrations up to twice as high as the concentrations of the training dataset. The $O_3$ was
actually better when being trained at the Downtown site and by using the RNN model could
predict concentrations up to 120 ppb. The NO$_2$ models were the worst performing models, even
the addition of electrochemical sensors was not able to improve the accuracy. The most effective
model, for measuring higher concentrations, was still the RNN with normalized targets however.

In this study DNN and especially RNN models have usually worked with higher accuracy than
MLR models across different devices and different sites. However, for measuring NO$_2$ the tested
metal oxide and electrochemical sensors do not seem to be able to achieve high accuracy across
multiple sites. Either newer sensors or methods for scrubbing out the cross-sensitive gases such
as $O_3$ are required before high confidence can be achieved.

7.2 Future directions

Low-cost devices offer many advantages to conventional pollution monitoring instruments
however many challenges remain. Throughout the previous chapters, it has been shown that NO
and CO electrochemical sensors are not very cross-sensitive to ambient levels of other pollutants
however they are very sensitive to environmental factors such as temperature and humidity.
Laboratory conditions may be used for calibration of these types of sensors, however, they are very sensitive to pressure so they must be calibrated and applied at the same pressure. These sensors must also be calibrated under varying temperature to the correct the values. If calibrating these sensors under ambient conditions they must be trained at a site with high ranges of concentration since they may under predict concentration otherwise. This effect can be mitigated by using a non-linear, time-dependent model like a RNN with a LSTM cell. Measuring NO$_2$ and O$_3$ is much more complicated with low-cost sensors since they experience high cross-sensitivity with other pollutants and each other. In the future more, complicated models such as RNN should be used for real-time measurements and visualization. This can be achieved by installing the relevant software such as TensorFlow on an online Linux server and loading models that were trained offline. Also, in real-time or at scheduled interviews, models may be trained or updated if a sensor is located next to a reference instrument. Unsupervised models, like clustering or Restricted Boltzmann Machine, may also be used in real-time to identify possible pollution sources in new environments.

In the future, electrochemical sensors should be used exclusively for measuring CO and NO. The main drawback is that these sensors are more expensive than metal oxide sensor arrays when including the ISB, which is required to process the signal from the sensor. However, if this board is developed in house the cost of the sensor will drop significantly. For NO$_2$ and O$_3$, the electrochemical sensor does not seem to perform as well as metal oxide sensors at the sites used for this study. Even when using metal oxide sensors the models are application specific. Therefore, it would be beneficial to train different models for difference applications. For instance, if this device is to be used for an indoor study, then the calibration model should be developed using the pollutant mixes and concentrations anticipated indoors. Conversely, if the device is to be used for outdoor urban measurements, then the models should be calibrated with ambient data. For studying health related impacts with these devices, data should be averaged more since the high time resolution data can be noisy and less precise.
References


[31] 7 million deaths linked to air pollution annually NIEHS: new WHO collaborating centre for environmental health Network to advance progress in children’s environmental health Launch of WHO International Scheme to Evaluate Household Water Treatment Techn, 2014.


[43] Alphasense, NO-B4 Nitric Oxide Sensor 4-electrode, 04/2015, 44 (n.d.) 0–1.


[71] Low Level NOx Data Sheet, Synkera Technologies, 09/2010, (n.d.).


