UNDERSTANDING OUR LOCAL ENVIRONMENT: DEVELOPING NOVEL APPROACHES TO QUANTIFY AND APPORTION AMBIENT VOCS WITH LOW-COST SENSORS

by

JACOB LINNEER THORSON

B.S., University of Wisconsin - Madison, 2013

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Master of Science
Department of Mechanical Engineering
2019

This thesis entitled:
Understanding Our Local Environment: Developing Novel Approaches to Quantify and Apportion Ambient VOCs with Low-Cost Sensors
written by Jacob Linneer Thorson
has been approved for the Department of Mechanical Engineering

Michael P. Hannigan, Ph.D.

Marina E. Vance, Ph.D.

Daven K. Henze, Ph.D.

Date ____________

The final copy of this thesis has been examined by the signatories, and we
find that both the content and the form meet acceptable presentation standards
of scholarly work in the above mentioned discipline.

Thorson, Jacob Linneer (M.S., Mechanical Engineering)

UNDERSTANDING OUR LOCAL ENVIRONMENT: DEVELOPING NOVEL APPROACHES TO QUANTIFY AND APPORTION AMBIENT VOCS WITH LOW-COST SENSORS

Thesis directed by Professor Michael P. Hannigan

ABSTRACT

In this dissertation, we demonstrate the application of low-cost air quality sensors to better understanding our local environment. Specifically, my work has focused on the application of arrays of low-cost sensors and methods of analysis that improve our ability to attribute local sources of volatile organic compounds (VOCs).

Low-cost sensors have been widely applied to the study of air quality at smaller spatial and temporal scales than was previously feasible. The research that is detailed in Chapter 2 built upon existing low-cost sensor research in order to develop an approach to both quantifying the concentrations of several compounds and also classifying the mixture based on the source that is likely to have emitted the detected gases. This research involved a chamber study where a large sensor array was exposed to complex gas mixtures that simulated realistic pollution sources. These data were used to validate the proposed methodology that involved a two-step process to accomplish the quantification and classification goals. The results of this approach show the feasibility of using low-cost sensors to directly estimate the effect of local sources of VOCs based on their chemical fingerprints.

ACKNOWLEDGEMENTS

I would like to offer my sincere thanks to my committee members, Nina Vance and Daven Henze, for their valuable time and insights. Thank you also to my advisor, Mike Hannigan, for his guidance, encouragement, and patience. Special thanks to Ashley Collier-Oxandale and Joanna Gordon Casey for their leadership and knowledge. Thank you also to David Pfotenhauer, Sarah Toth, Evan Coffey, and the rest of the Hannigan research group for welcoming me and helping to make this research possible. Additional thanks to Gustavo Aguirre Jr. and Sue Chiang for providing an opportunity to work directly with communities affected by this research. Last and certainly not least, thank you to my wonderful wife, Bethany Thorson, for supporting me always.

This research is supported in part by the AirWaterGas Sustainability Research Network funded by the National Science Foundation under Grant No. CBET-1240584. Any opinion, findings, and conclusions or recommendations expressed in this dissertation are those of the author and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

APPENDIX

LIST OF TABLES

CHAPTER 1:

CHAPTER 2:

APPENDIX A:

APPENDIX B:

LIST OF FIGURES

CHAPTER 1:

CHAPTER 2:

APPENDIX C:

CHAPTER I

INTRODUCTION

1. Overview

This chapter provides an introduction to volatile organic compounds (VOCs) and the importance of understanding their emissions and distribution through our communities. The introduction starts by reviewing some of the most common sources affecting air quality across the United States in order to review the current understanding of differences within and between those sources in regards to the VOCs that they emit. A specific eye is then turned toward the methodologies that are currently used for the detection and quantification of VOCs. After understanding how current methods can provide information about VOC distribution and emissions, I will provide an overview of the next generation of air quality monitors, with a specific eye to the low-cost sensor technologies that are currently commercially available. Finally, I will provide a summarize the knowledge gaps that the work presented in Chapter 2 addresses.

2. Importance of VOCs

More than 100 million Americans live in communities where air pollution standards are annually exceeded (USGCRP, 2018). Among the compounds thought to affect human health, either directly and or indirectly, are a class of compounds referred to as VOCs. VOCs are a class of compounds that broadly refer to any atmospheric organic exclusive of CO and CO_2 (Seinfeld & Pandis, 2016) and are emitted from everything from heavy industry to personal care products. VOCs are often subdivided between methane and nonmethane hydrocarbons (NMHCs) because of difference in the typical background concentrations. VOCs are also sometimes subdivided by their volatility, ranging from semi-volatile organic compounds (SVOCs) to very volatile compounds (VVOCs). In this paper, references to VOCs will include all of the above unless otherwise noted.

The emission and transport of VOCs is important because of the health and environmental effects of directly emitted VOCs and their secondary products. Many of the compounds included on the US EPA list of hazardous air pollutants are VOCs (OAR US EPA, 2015b). Human health is affected by both acute and chronic exposures to these compounds. The US EPA maintains the Integrated Risk Information System (IRIS) with the goal of characterizing the toxicological risks associated with many compounds of interest. For example, chronic exposure to benzene, a VOC emitted from sources including gasoline vapor and tobacco smoke, has been linked to leukemia and other adverse health effects from workplace exposures at 8-hour concentrations as low as 7.6 ppm (ORD US EPA, 2003). The monitoring and emission of several species of VOCs are regulated by the US EPA and the EU as well as other regulatory agencies because of demonstrated health effects. The European Parliament has also mandated that ambient VOC concentrations be monitored (Directive 2008/50/EC, 2008). The emissions of VOCs from household products are also regulated because of the effects of high concentrations of VOCs experienced in indoor environments with inadequate ventilation as well as their contributions to atmospheric VOC concentrations.

In addition to known and suspected direct health effects, VOCs have been shown to contribute to the formation of particulate matter (PM) and ozone (Khordakova, 2017; Ziemann & Atkinson, 2012). Both of these are "criteria pollutants" and are regulated because of known associated health effects, especially related to lung function (US EPA, 2006). VOCs also react in the atmosphere to contribute to the production of other known air toxics like carbon monoxide, formaldehyde, and other reactive compounds (de Gouw et al., 2018).

3. Origins and Fate of VOCs

3.1. VOC Sources

VOCs are produced and emitted from an incredible range of sources including evaporating solvents in personal care products (Coggon et al., 2018), oil and gas infrastructure (Warneke et al., 2014), wildfires (Yokelson et al., 2013), and vehicle emissions (Westerholm & Egebäck, 1994). Anthropogenic sources make up only about 10% of total global VOC emissions (Seinfeld & Pandis, 2016), although this can vary significantly at smaller scales. The contribution of sources of VOCs as estimated by the 2014 National Emissions Inventory (NEI) Report produced by the US EPA is illustrated at a national scale in Figure 1 and at smaller scales in Figure 2. As is apparent when comparing these figures, the contributions of different sources vary when viewed at national, state, or local scales, even when considering major classes of sources on an annually averaged basis.

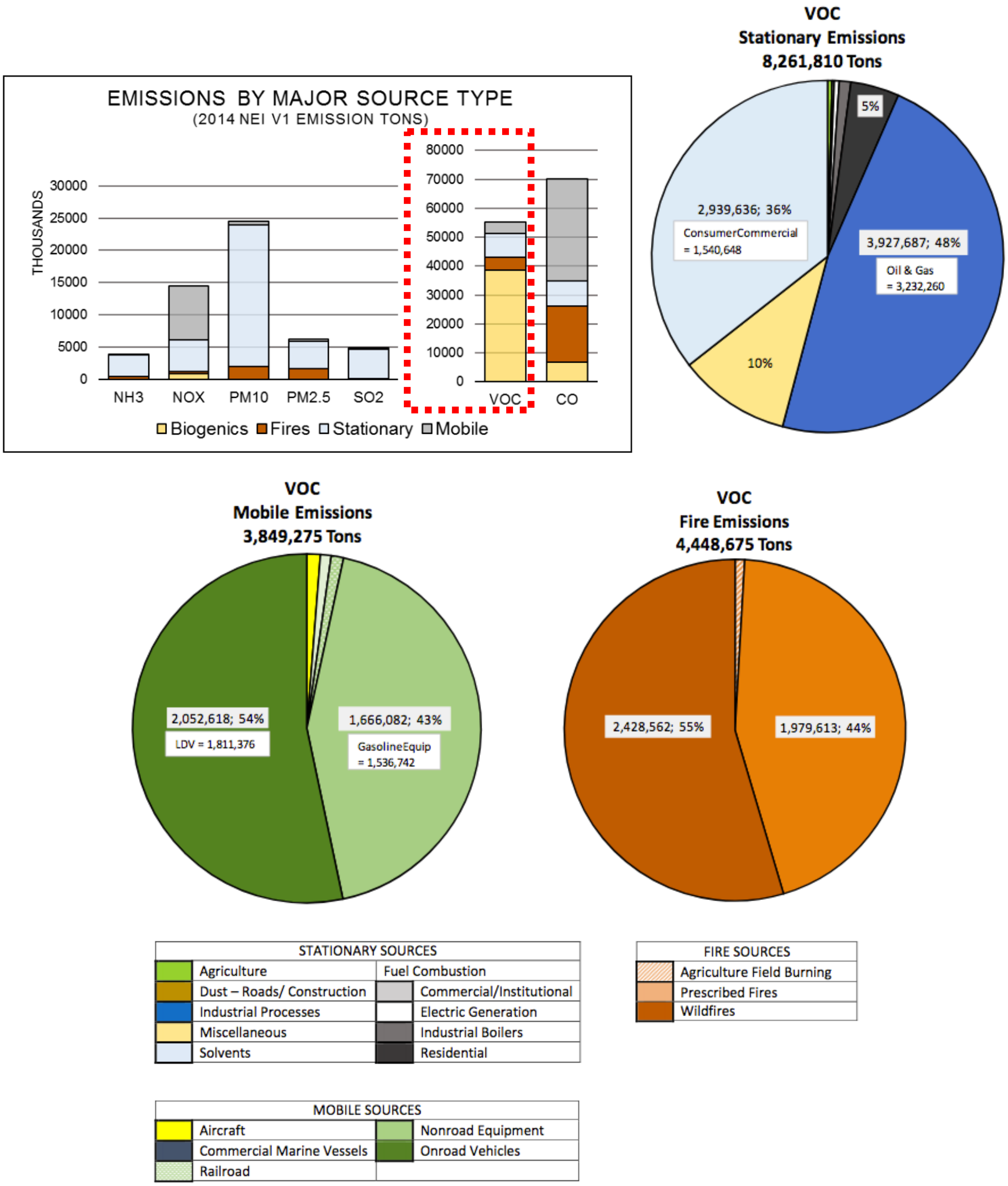


Figure 1. Contributions of different source types to the annual emissions of VOCs within the US as reported in the national emissions inventory for 2014 (OAR US EPA, 2016a). Clockwise from top left: the contributions of broad categories are shown to multiple pollutants including VOCs, which are highlighted with a dashed red box; stationary VOC sources are subdivided by sectors; fire emissions are subdivided by sectors; mobile sources are subdivided by sectors.

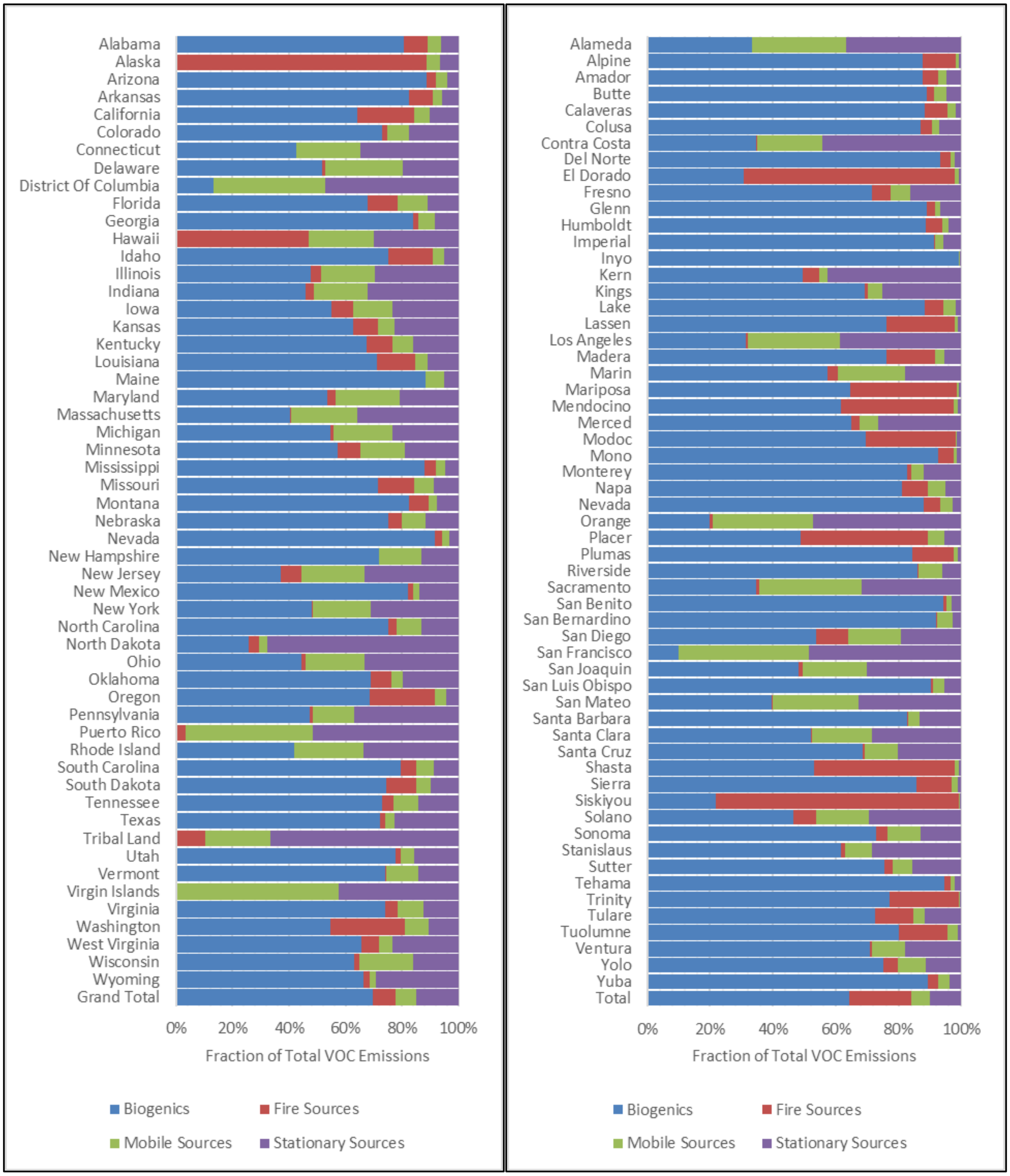


Figure 2. Fraction of VOCs attributed to each class of source for each state (top) and each county within the state of California (right). The contribution of source types to the total VOC emissions from the state or county varies at not just the national level, but also at state and county levels. Data from (OAR US EPA, 2016a). Note that some states and counties did not report emissions values for all source types.

VOCs are studied in many different ways that will be discussed later because of their importance to human health and the environment depending on the specific compounds that are emitted as well as the temporal and spatial distribution of the emissions. Many different approaches to quantifying VOC emissions are compiled to create emissions inventories that estimate the typical contributions of sources on scales that typically range from global (Saunois et al., 2016) to national and state-level (OAR US EPA, 2016a). While these inventories are important tools for developing regulations and making some decisions, they often lump VOCs into a single category, neglecting the incredible variety of VOC species, and the different impacts that individual species have on atmospheric chemistry and human health (Ditto et al., 2018). Although there is significant variation within the different source types, it is possible to make some general statements about typical chemical components of different sources, something that is important when attempting to identify source emissions based on measurements of different gas concentrations alone. Understanding the variation in chemical composition within and between VOC sources is useful for understanding the chemical properties of those emissions – an important consideration when attempting to measure them and understand their transport and fate.

3.1.1. Biomass Burning

The single largest source of VOCs in the United States as identified in the NEI was fires. This category includes both wildfires, prescribed burns, and field burning, although US emissions are dominated by the first two categories of fires, in regards to VOC emissions (OAR US EPA, 2016a). Biomass burning is also one of the largest anthropogenic emitters of VOCs in across the planet (Akagi et al., 2011). These VOCs are emitted in both gaseous form and sorbed on particulates, making their total quantification a challenge.

The emissions of VOCs from fires depend heavily on the combustion state and fuel source for the fire (Gilman et al., 2015). The VOC emissions from biomass burning typically include highly oxygenated compounds as well as many nitrogen containing compounds (Yokelson et al., 2013). These compounds include everything from levoglucosan, a relatively harmless product of pyrolysis that is a common tracer for low-temperature combustion (Lai et al., 2019) to hydrogen cyanide and formaldehyde, both compounds with well-established health effects (Yokelson et al., 2013). The 2014 NEI also found that biomass burning was the dominant source of formaldehyde emissions in the US (OAR US EPA, 2016a).

Although not all of these compounds have known health effects, several studies have found that the emissions of CO, NO_x, and VOCs from wildfires can directly contribute to increased ozone production in regions downwind of the fires (Baker et al., 2016; Pfister, Wiedinmyer, & Emmons, 2008). Pfister et al. showed that the NO_x produced by the fires was consumed quickly through reactions with other combustion products, producing peroxyacetyl nitrate (PAN) and other nitrogen containing compounds, indicating that the contribution to ozone production was likely due to the contributions of CO and VOCs (Pfister et al., 2008).

Fires as sources of VOCs are important for areas of the US, specifically the west and south, where wildfires are common, and in areas of the Midwest where the burning of crop residues can lead to large, localized emissions (NASA, 2011).

3.1.2. Mobile Sources

Mobile sources, like vehicle traffic and other internal combustion engines, are the second largest contributor of VOCs identified by the NEI (OAR US EPA, 2016a). Within the “mobile sources” category identified by the NEI, VOC emissions were dominated by light duty vehicles and other nonroad gasoline equipment. Mobile sources are also important on a global scale, although typical mobile emissions can vary significantly between countries (Hoesly et al., 2018). Recently, the European Environmental Agency (EEA) also estimated that road transportation exhaust contributes about 7% of total VOC emissions from member countries, compared to a roughly 12% contribution in the US as estimated in the NEI (European Environmental Agency, 2017; OAR US EPA, 2016a).

The composition and amount of VOCs emitted from a given vehicle vary significantly depending on the vehicle design, operating conditions, fuel type, and emissions control technologies (Faiz, 1996). Vehicle traffic emits VOCs directly as both a gas and as a major component of emitted particulate matter (Westerholm & Egebäck, 1994). These VOC emissions may be especially troublesome as they are co-emitted with NO_x, which is highly reactive and, in the presence of CO and VOCs, can lead to significantly increased ozone production. Traffic sources are both highly variable and intertwined with communities such that extremely dense networks of continuously measuring sensors (~40-100m grid) would be necessary to get accurate estimate of exposure to traffic related air quality effects (Batterman, Chambliss, & Isakov, 2014). VOCs are also typically products of incomplete combustion and as such contain many oxygenated VOCs (OVOCs) that are partially oxidized forms of the original fuel mixtures (Westerholm & Egebäck, 1994). Because the fossil fuel products that typically make up the bulk of the fuel contain long-chain alkanes, alkenes, and polycyclic aromatic hydrocarbons (PAHs), forms of those compounds often appear in vehicle exhaust (Westerholm & Egebäck, 1994).

Although conclusions may be drawn about typical vehicle sources, there is variation between vehicles, with high emitting vehicles significantly contributing to the total emissions (Tan, Lipsky, Saleh, Robinson, & Presto, 2014). This means that that air quality can be significantly affected by the specific vehicles operating in a given locale.

3.1.3. Fossil Fuel Infrastructure

With the boom in tight and shale gas production in the United States, fossil fuel production has grown to affect areas both historically known for this production like Texas and Southern California, as well as new regions like Colorado, the Dakotas (US EIA, 2015). It can be seen in Figure 3, that the production of fossil fuels is highly concentrated in some regions, while nearly absent from others.

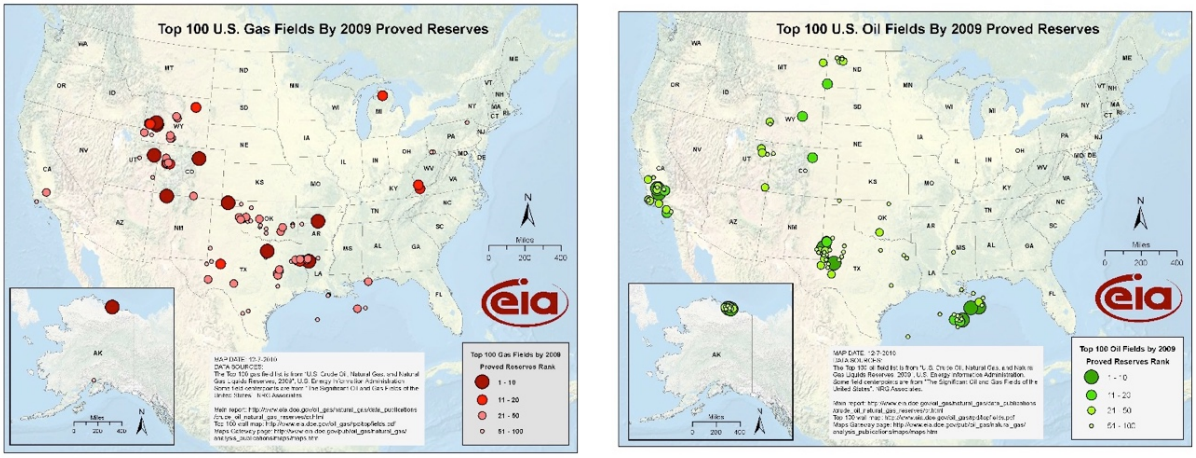


Figure 3. The 100 largest natural gas (left) and oil (right) fields as ranked by proved reserves span across regions of the South, West, and parts of the Midwest (US EIA, 2015).

Although production is limited to certain areas of the US, the natural gas infrastructure in some cities can be over 100 years old and prone to leakage contributes significantly to methane emissions on a national and local scale (Gallagher et al., 2015; Heath, Warner, Steinberg, & Brandt, 2015; Phillips et al., 2013). In fact, the US EPA estimates that natural gas systems are the second largest emission source of methane emissions, just behind those emissions from enteric fermentation (OA US EPA, 2018). Recent field measurements of methane emissions from the oil and gas supply chain have produced methane estimates that are as much as 60% higher than the emission inventory kept by the US EPA (Alvarez et al., 2018). In this study, Alvarez, et al. suggest that the reason behind this discrepancy is the nature of most other studies that are conducted once and may not capture periodic or sporadic events that produce extreme emission conditions but only for a brief period of time.

The emissions of specific compounds can once again vary significantly in composition and magnitude depending on the region and specific activity that is responsible for the given emissions. For example, one common emission path is through open "thief hatches" or other vents that become stuck open and continuously leak vaporized light hydrocarbons (Alvarez et al., 2018), whereas emissions from natural gas infrastructure as mentioned earlier would be dominated by methane and other light alkanes. While I was working at the SwRI, my collaborators and I found that even processed gas available at natural gas vehicle (NGV) filling stations varied between regions, despite significant upstream processing (George, Poerner, Ridens, & Thorson, 2014).

In general, nonmethane alkanes are often used as indicators of fossil fuel production emissions when studying emissions on a regional scale (Katzenstein, Doezema, Simpson, Blake, & Rowland, 2003). More detailed studies of ambient compounds attributable to oil and gas production have found a wide mix of alkanes, alkenes, aromatics, and other hydrocarbons (Goldan et al., 2004; Harley, Hannigan, & Cass, 1992). This qualitatively matched the findings of Ryerson et al. who found that vaporized compounds from spilled crude oil was dominated by straight and cyclic isomers of alkanes of a wide range of carbon number (Ryerson et al., 2011).

3.1.4. Industrial and Household Solvents

The solvents that make up major components of paints, lubricants, cleaners, and a variety of other chemical products are often some combination of VOCs. They are estimated to be responsible for emissions comparable to those attributed to all light duty vehicles (OAR US EPA, 2016a). In some areas, solvents are responsible for significantly more VOC emissions thanks in part to recent regulations that were targeted at reducing mobile VOC emissions (McDonald et al., 2018).

The chemical feedstocks that go on to produce many of these products are largely derived from fossil hydrocarbons (McDonald et al., 2018). Solvents and the products that they are part of include not only those hydrocarbons but include a wide range of compounds depending on their application. For example, acetone is a common chemical marker for coatings-related emissions (McDonald et al., 2018; Warneke et al., 2014). Cyclic volatile methyl siloxanes (cVMS), on the other hand, are a class of organic compound that is popular in the cosmetics and personal care products industry due to their volatility and low odor (Wang, Moody, Koniecki, & Zhu, 2009). These compounds were found by Coggon et al. to have distinct diurnal patterns above cities that corresponded with morning applications of deodorant and other products around 6-7 AM (Coggon et al., 2018).

It is often difficult to identify compounds associated with commercial and industrial products because they often include OVOCs that are both highly reactive and potentially produced by the atmospheric reactions of other emitted compounds (Niedojadlo, Becker, Kurtenbach, & Wiesen, 2007).

3.1.5. Agriculture and Waste Treatment

The last grouping of VOC sources to be discussed here are agriculture and waste treatment, which may be roughly considered together because a large portion of VOC emissions related to both of these sources involve anaerobic digestion and the subsequent products. This digestion occurs in the digestive tracts of animals, in the soil of fields (Linquist, Groenigen, Adviento-Borbe, Pittelkow, & Kessel, 2012), and in the piping and processing tanks associated with wastewater treatment (Guisasola, Sharma, Keller, & Yuan, 2009). Depending on the specific identity of any of these sources, the quantity of emissions can vary by orders of magnitude and at timescales of seasons to days.

For example, emissions from enteric digestion and soil bacteria associated with agriculture are typically associated with the production of methane and nitrous oxide (N_2O), two powerful contributors to global warming (Linquist et al., 2012; Pétron et al., 2012). Howard and colleagues also found that the feedstock for animals may directly emit reactive organic compounds (ROGs) at levels that affect ozone production in California's central valley (Howard et al., 2010). In addition to the variety of agricultural VOC sources, Mihelcic found that wastewater treatment facilities emitted significant amounts of VOCs, many of which were chlorine or sulfur-containing compounds (Mihelcic, Baillod, Crittenden, & Rogers, 1993).

3.2. VOC Sinks

VOCs may be removed from the atmosphere in a variety of ways. Some of these processes create secondary products that are themselves cause for concern. For example, VOCs are known to participate in the HO_x-NO_x cycle that is a major driver of ozone production in the troposphere. For example, McDuffie found that VOC emissions in the Front Range region of Colorado, VOCs contributed to local ozone production (McDuffie et al., 2016). Wu et al. also showed that changes to meteorology caused by climate change will counteract some of the predicted declines in anthropogenic VOC emissions in terms of episodic ozone events, making their detection and control even more vital going forward (Wu et al., 2008).

During daylight hours, one of the most common pathways for atmospheric reactions of VOCs begins when a compound is oxidized by reactions with OH radicals as illustrated in Figure 4, although reactions with NO_2, NO_3, and O_3 can be competitive depending on local conditions. After that initial reaction, the VOC participates in the HO_x-NO_x cycle by recycling NO to NO_2 where it may photolyze, contributing a free oxygen and subsequently ozone. The product from the original VOC compound after this reaction is often less volatile and therefore more likely to partition into particle phases (Ziemann & Atkinson, 2012). The acidic or basic chemical environments that develop within these particulates may then contribute to further reactions, either allowing the new compound to oligomerize, further reducing its volatility, or to further decompose, possibly allowing it to return to the vapor phase (Ziemann & Atkinson, 2012). These reactions can happen on a timescale ranging from the near-instantaneous to days or years, depending on the specific chemical compound and environmental factors like temperature, solar radiation, and the presence of other compounds.

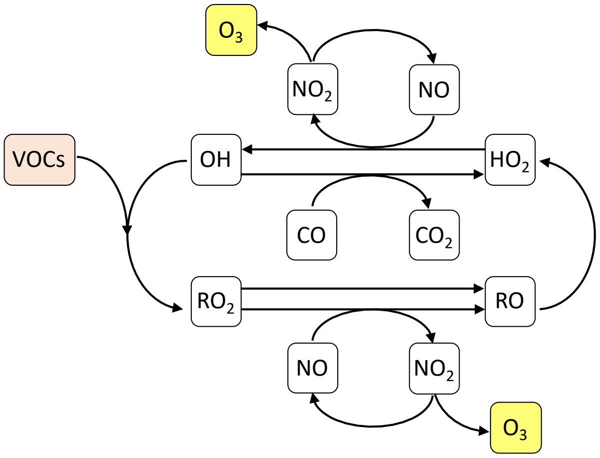


Figure 4. A simplified chemical cycle illustrating how VOCs can participate in the daytime HO_x-NO_x cycle and contribute to the production of ozone. Adapted from (Clemitshaw, 2004).

4. Approaches to Understanding Our Local Environment

In order to understand the incredible variation in VOC emissions that has only been superficially described above, many approaches have been developed to improve our understanding of VOC emissions and transport. Leaving aside other differences, these monitoring techniques can generally be thought of as varying on two axes that represent the resolution in both space and time as illustrated in Table 1. In the time domain, one can broadly say that monitoring technologies consist of either continuous, ongoing measurements or of measurement campaigns spanning brief periods of time. In the spatial domain, studies may capture different degrees of resolution that may be categorized as regional, i.e. on a scale of tens to hundreds of kilometers, or as local measurements that are only separated by tens to hundreds of meters.

Table 1. Broad categorization of paradigms for understanding VOC sources

		Spatial Resolution	
		Low (Regional)	High (Local)
Time Resolution	Low (Infrequent)	Compilation of Emissions Inventories	Measurement Campaigns
	High (Continuous)	Regulatory Monitoring	Satellites, Low-Cost Sensors

4.1. *Regional and Infrequent*

In the top left region of Table 1 are methods for understanding VOC sources that produce information with both low spatial and time resolutions. An example of these might be the compilation of emissions inventories, which are actually a synthesis of many individual information sources. Emissions inventories attempt to describe the emissions from many different sources over a large region by synthesizing information from studies ranging from continuous monitors to brief on-site measurements. These data are applied to estimate the long-term averages of emission factors, the locations of those sources, and the activity level of each source (OAR US EPA, 2015a). These are useful for understanding relative contributions of different sources for many different pollutants on an annually averaged basis. Typical uses for the data that is produced by an emissions inventory are regulatory actions, where it is important to understand what types of sources most impact the production of a specific compound. Another example application are scientific studies where it is important to understand whether current research is able to account for all of the compounds measured in our atmosphere. If not, the shortfalls may inform the design of future research programs.

4.2. Infrequent but Highly Spatially Resolved

Moving to the top right quadrant of Table 1, are approaches that allow for high spatial resolution but are infrequent, and include methods like one-time field campaigns. Detailed measurement campaigns like the DISCOVER-AQ in Colorado's Front Range are a good example of this approach (Halliday et al., 2016). In that study, a series of measurements were taken across the front range using a combination of aerial flyovers and ground measurements with the goal of characterizing benzene and other VOC emissions in both urban areas and rural areas that are in or nearby to concentrated oil and gas production. This approach produced a high-resolution map of specific VOC concentrations as collected over the course of two months in 2014. These types of measurement campaigns can provide high accuracy, high resolution snapshots in time that are useful for answering research questions. In the case of DISCOVER-AQ, researchers sought to determine the impact that oil and gas production had on the local airshed. At the conclusion of the study, they were able to determine that local oil and gas production was the primary source of benzene in the Denver area (Halliday et al., 2016). However, these measurement campaigns are quite expensive, requiring multiple research grade instruments and hundreds of hours of skilled technical time. This cost makes them infeasible to conduct frequently, let alone on a continuous basis, and they are therefore unable to capture sporadic, unpredictable events like accidental emissions that are suspected to be responsible for the current underestimation of emissions of methane from oil and gas production (Alvarez et al., 2018).

4.3. Regionally Spaced but Frequent

The bottom left of Table 1 represents the opposite of the previous methodology – methods that provide continuous measurements at very low resolution. A prominent example of this methodology are the networks operated by state and national regulatory agencies that monitor compounds that are important for regulatory decisions. These are less common in the United States for VOCs, which are not themselves a criteria pollutant. One network that does monitor VOC concentrations across the US is the network of Photochemical Assessment Monitoring Stations (PAMS) which operates approximately 145 monitors collect samples on the order of hours to days for specific VOCs with the primary goal of determining their effect on ozone production (OAR US EPA, 2016b). State level entities also commonly use this approach, for example, the network of monitors operated by the California Air Resources Board (CARB) is illustrated in Figure 5 and reports the concentrations total nonmethane hydrocarbons (TNMHC) on an hourly basis. Regulatory networks like those mentioned produce continuous measurements, typically reported as hourly averages of all VOCs, methane and NMHCs, or compounds grouped by chemical attributes. These measurement stations include regulatory grade equipment that is expensive to purchase and operate, so they are placed with the goal of measuring typical concentrations affecting a given area or population. Because the goal when establishing these networks is typically to quantify the air affecting the most people, they are placed tens to hundreds of kilometers apart. These networks provide information to regulators about important pollutants to, for example, determine if a region is meeting air quality standards. They are not, however, able to quantify the variation in effects of pollutants that themselves vary on much smaller scales. These including NO_x, ozone, and VOCs (Batterman et al., 2014; Cheadle et al., 2017; Coggon et al., 2018).

Figure 5. Approximate locations of the 19 monitors operated by CARB that are reporting measurements of total nonmethane hydrocarbon (TNMHC) concentrations. Data from (CARB, 2017) overlaid on a map from (Dedering, 2010).

The three methodologies discussed above each have their place in answering scientific, research, and regulatory questions, but each also sacrifices time or spatial resolution to be able to produce accurate, affordable measurements in the dimension of interest. Many of the products from these approaches also reduce their exploration to concentrations of total VOCs, which as discussed above removes the wide variation in the makeup of VOCs emitted by different sources. Commonly used statistical analyses like principal component analysis (PCA), kriging, back-propagation, and others may help to address some of these gaps, but all include their own assumptions that reduce accuracy.

4.4. Next Generation Technologies

The newest generation of low-cost sensors attempts to capture continuous, high spatial resolution measurements with the goal of developing an approach that fills the lower right quadrant. This increased resolution is typically at the cost of measurement accuracy in the form of increased bias and random signal noise. One approach that will not be discussed here is the newest generation of Earth-facing satellites. These show the promise of sub-kilometer resolutions with re-visit times on the order of days to weeks (ESA, 2017). These can be quite accurate in their measurements of methane concentrations, but they are limited in their ability to determine the vertical distribution of the measured compounds.

In the past approximately two decades, "low-cost" environmental sensor technologies have matured, and the associated costs have decreased to the point that they are becoming a popular option for researchers and community members looking to understand their local environment. The Hannigan Lab at CU Boulder developed several platforms to incorporate some of these low-cost sensors to study everything from cookstove efficiency to the distribution of methane across a community in downtown Los Angeles. The platform used in this study, the Y-Pod has a unit cost that is approximately 1% of the cost of a research grade instrument for the same compound and produces continuous measurements. The relatively low cost of these sensors makes it feasible to distribute them across a community of interest and investigate the relative effect of different VOC sources on that community (A. Collier-Oxandale et al., 2018; Shamasunder et al., 2018). These distributed sensors can be used with many of the same statistical analysis techniques applied to more expensive, research grade instruments, albeit with generally lower accuracies. We discussed the pitfalls and general guidelines for producing and communicating data using low-cost sensors after meeting with industry and community leaders (Clements et al., 2017). In that meeting, stakeholders discussed the many questions yet to be answered about the appropriate ways to operate and understand low-cost sensors but also the promise of understanding air quality at scales smaller than was previously feasible. Some low-cost air quality sensor technologies will be briefly discussed next in order to illustrate the diversity of technologies being investigated in this space.

4.4.1. Metal Oxide

Metal oxide sensors as a technology for the measurement and quantification of gaseous pollutants has been a subject of study since the discovery of semiconductor's sensitivity to reacting gases in the early 1950s (Brattain & Bardeen, 1953). Metal oxide sensors are a popular option in the low-cost sensing community because of their especially low cost (~$10 USD), small size and low weight, and relative robustness when exposed to the environment. The operation of these sensors involves bringing a metal oxide sensing media to a few hundred degrees Celsius in an oxygen containing atmosphere, typically using a printed platinum heater that is separated from the sensor material with a thin layer of alumina. In the presence of a reducing gas, the concentration of surface-adsorbed oxygen species decreases as that gas reacts with the oxygen at the surface of the sensor. This reaction releases electrons into the conductance band, affecting the measured conductance of the sensor (Windischmann & Mark, 1979).

As the understanding of these sensors progressed, studies began to investigate the possibility of quantifying and identifying VOCs with metal oxide sensors using various operation and manufacturing schemes. For example, some have looked at the time response of sensors to changing temperature as a way to improve sensitivity to VOCs and/or to allow for some level of identification of the mixture of interest (Bastuck, Reimringer, Conrad, & Schütze, 2016; Schultealbert, Baur, Schütze, & Sauerwald, 2018; Sears, Colbow, & Consadori, 1989; Vergara et al., 2005). Manufacturers have also developed metal oxide sensors that are designed to specifically react to VOCs, through a combination of intentional design of the heater temperature and selection of the base metal oxide and doping compounds.

Korotcenkov et al. discussed the most common issue affecting the operation of metal oxide sensors: the drift in both baseline signal and sensitivity over time (Korotcenkov & Cho, 2011). The rate at which these parameters drift may vary on the order of days to months depending on the processes responsible for the drift. They found that perceived drift may be attributed to a range of causes including physical changes at the atomic level, poisoning from exposure to other gases, and changes in the local environmental parameters including temperature, humidity, and concentrations of confounding species.

4.4.2. Electrochemical

Electrochemical sensors are another technology with an extensive history of use as sensors for gases and operate on much the same principle as the fuel cells that are used to power next generation vehicles, albeit at a much smaller scale. In these sensors there are two primary electrodes, the working and counter electrodes, that are kept in electrical contact via an electrolyte solution. Both of these electrodes have a catalytic surface that is exposed to the ambient environment. At the working electrode, this surface is designed to react with a reducing gas (CO, NO, SO_2, select VOCs, etc.) or oxidizing gas (NO_2, O_3, etc.). When the reaction of the target gas occurs at the surface of the sensor, a small current is generated through the electrolyte causing the opposite reaction at the counter electrode (often the reduction of oxygen when sensing reducing gases). When operated correctly this current is approximately proportional to the concentration of the gas of interest (Alphasense Ltd., n.d., p. 104). However, because the sensing surfaces are exposed to the ambient environment, they may be affected by other gases with similar properties and may even be “poisoned” if gases are present that adsorb onto the catalytic surface. For example, Alphasense has reported that different compounds were responsible for poisoning the catalytic surfaces with different target gases (Alphasense Ltd., n.d.). Electrochemical sensors are also sensitive to changes in humidity because the electrolyte solution is exposed to the environment, and they are known to eventually dry out and lose functionality after some time, again depending on local conditions (Alphasense Ltd., n.d.).

4.4.3. Optical Methods

Sensors based on nondispersive infrared (NDIR) and photoionization detection (PID) are two measurement technologies that can range from low-cost sensors on the order of tens of USD to research grade instruments that are tens or hundreds of thousands of dollars. Both of these sensor types operate by interrogating a gas sample with infrared or high-energy, typically ultraviolet, light for NDIR and PID, respectively (Aleixandre & Gerboles, 2012). For NDIR sensors, an optical detector opposite the gas sample from the infrared source detects the reduction in signal associated with the presence of the target gas. The precise wavelength of the sensor can be tuned to attempt some level of specificity depending on the optical properties of the target gas and likely interfering gases. PID sensors instead operate by detecting the flow of ions produced by the photoionization of target molecules by an ultraviolet light source. This can be targeted somewhat more specifically by tuning the ultraviolet source to ionize VOC molecules. By careful selection of the photon energy produced by the UV lamp, an approximate measurement of total hydrocarbons can be produced because of their relatively low ionization energy when compared to other major components of air. However, the response of these sensors depend on the specific VOCs, and they does not allow for speciation without a coupled gas chromatography column or another separation method. Both of these technologies have the benefit that they operate on optical principles and may therefore react on a nearly instantaneous basis to changes in concentration within their optical path.

5. Knowledge Gaps that will be Addressed

Understanding which VOC sources are affecting air quality at a human scale is an important step to addressing sources that may be missed by current regulatory practices. Addressing localized sources is an important step toward addressing disparities in air quality that drive unequal health burdens (Martenies, Milando, Williams, & Batterman, 2017; Shamasunder et al., 2018). In the past, detecting specific sources using low-cost sensors has been accomplished in a few ways. The foremost uses the advantage of a low unit cost to create dense networks that can be combined with wind direction information to determine the likely location from which a compound was emitted (Ashley Collier-Oxandale, Coffey, Thorson, Johnston, & Hannigan, 2018; Heimann et al., 2015; Mead et al., 2013). A similar approach is the use of mobile sensors, leveraging the low-weight nature of sensors to create a map of concentrations (Penza, Suriano, Pfister, Prato, & Cassano, 2017). A third, less targeted approach, is the removal of a baseline measurement that is either determined using upwind measurements or through forms of timeseries analysis with some assumptions about the temporal variation of far-field versus nearby sources (Klems, Pennington, Zordan, & Johnston, 2010; Popoola et al., 2018; Sabaliauskas, Jeong, Yao, & Evans, 2014).

In most studies that attempt to identify pollutant sources using low-cost sensors, there is some assumption that a type of source is predominantly responsible for the emission of the measured compound(s), and that therefore measuring a few compounds can map the spatial effects of that source on the local air quality. Removing that assumption by improving the ability of low-cost sensor systems to identify the source type of pollutants would make those results more parsimonious and more likely to produce accurate conclusions. In this study, we demonstrated a system that can provide information about gaseous concentrations and also identify the likely source type at a cost point that is feasible for deployment at a sub-kilometer scale (roughly $1000 USD).

One common question encountered during field deployments of low-cost sensors is the sensitivity of specific sensors to confounding gases and their target gases. This is an important feature of the sensors to understand because of the diversity of compounds that are produced by different sources. However, manufacturers often do not provide detailed information about the effects of confounding gases and also do not typically make claims about sensor response at concentrations low enough to be relevant for ambient air monitoring. By placing a large array of low-cost sensors in a controlled environment and exposing them to complex, environmentally-relevant gas mixtures generated data that can answer questions about the sensitivity of both individual sensors as well as the full array. Another benefit of the study was the ability to vary gas concentrations and environmental conditions independently, an approach that provided information that is not possible to collect during a field deployment. This information will inform the current understanding of which sensors are useful when attempting to study a specific compound in complex environments. Finally, we will study whether an array of sensors can be used to algorithmically identify likely sources of VOCs.

6. REFERENCES

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., … Wennberg, P. O. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, *11*(9), 4039–4072. https://doi.org/10.5194/acp-11-4039-2011

Aleixandre, M., & Gerboles, M. (2012). Review of Small Commercial Sensors for Indicative Monitoring of Ambient Gas. *Chemical Engineering Transactions*, *30*.

Alphasense Ltd. (n.d.). AAN 104: How Electrochemical Gas Sensors Work. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_104.pdf

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., … Hamburg, S. P. (2018). Assessment of methane emissions from the U.S. oil and gas supply chain. *Science*, eaar7204. https://doi.org/10.1126/science.aar7204

Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., … Pierce, T. (2016). Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches. *Atmospheric Environment*, *140*, 539–554. https://doi.org/10.1016/j.atmosenv.2016.06.032

Bastuck, M., Reimringer, W., Conrad, T., & Schütze, A. (2016). Dynamic Multi-sensor Operation and Read-out for Highly Selective Gas Sensor Systems. *Procedia Engineering*, *168*(Supplement C), 1685–1688. https://doi.org/10.1016/j.proeng.2016.11.490

Batterman, S., Chambliss, S., & Isakov, V. (2014). Spatial resolution requirements for traffic-related air pollutant exposure evaluations. *Atmospheric Environment*, *94*, 518–528. https://doi.org/10.1016/j.atmosenv.2014.05.065

Brattain, W. H., & Bardeen, J. (1953). Surface Properties of Germanium. *Bell System Technical Journal*, *32*(1), 1–41. https://doi.org/10.1002/j.1538-7305.1953.tb01420.x

CARB. (2017, July 25). Sites Measuring Total Hydrocarbons. Retrieved December 20, 2018, from https://www.arb.ca.gov/qaweb/sitelist_generator.php

Cheadle, L., Deanes, L., Sadighi, K., Casey, J. G., Collier-Oxandale, A., & Hannigan, M. (2017). Quantifying Neighborhood-Scale Spatial Variations of Ozone at Open Space and Urban Sites in Boulder, Colorado Using Low-Cost Sensor

Technology. *Sensors; Basel*, *17*(9), 2072. http://dx.doi.org.colorado.idm.oclc.org/10.3390/s17092072

Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., … Hannigan, M. (2017). Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). *Sensors*, *17*(11), 2478. https://doi.org/10.3390/s17112478

Clemitshaw, K. (2004). A Review of Instrumentation and Measurement Techniques for Ground-Based and Airborne Field Studies of Gas-Phase Tropospheric Chemistry. *Critical Reviews in Environmental Science and Technology*, *34*(1), 1–108. https://doi.org/10.1080/10643380490265117

Coggon, M. M., McDonald, B. C., Vlasenko, A., Veres, P. R., Bernard, F., Koss, A. R., … de Gouw, J. A. (2018). Diurnal Variability and Emission Pattern of Decamethylcyclopentasiloxane (D5) from the Application of Personal Care Products in Two North American Cities. *Environmental Science & Technology*, *52*(10), 5610–5618. https://doi.org/10.1021/acs.est.8b00506

Collier-Oxandale, A., Hannigan, M. P., Casey, J. G., Piedrahita, R., Ortega, J., Halliday, H., & Johnston, J. (2018). Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. *Atmos. Meas. Tech. Discuss.*, *2018*, 1–35. https://doi.org/10.5194/amt-2017-421

Collier-Oxandale, Ashley, Coffey, E., Thorson, J., Johnston, J., & Hannigan, M. (2018). Comparing Building and Neighborhood-Scale Variability of CO2 and O3 to Inform Deployment Considerations for Low-Cost Sensor System Use. *Sensors (Basel, Switzerland)*, *18*(5). https://doi.org/10.3390/s18051349

de Gouw, J. A., Gilman, J. B., Kim, S.-W., Alvarez, S. L., Dusanter, S., Graus, M., … Young, C. J. (2018). Chemistry of Volatile Organic Compounds in the Los Angeles Basin: Formation of Oxygenated Compounds and Determination of Emission Ratios. *Journal of Geophysical Research: Atmospheres*, *123*(4), 2298–2319. https://doi.org/10.1002/2017JD027976

Dedering, U. (2010). *Location map of the USA (without Hawaii and Alaska).* Retrieved from https://commons.wikimedia.org/wiki/File:Usa_edcp_location_map.svg

Directive 2008/50/EC (2008). Retrieved from http://data.europa.eu/eli/dir/2008/50/oj

Ditto, J. C., Barnes, E. B., Khare, P., Takeuchi, M., Joo, T., Bui, A. A. T., … Gentner, D. R. (2018). An omnipresent diversity and variability in the chemical composition of atmospheric functionalized organic aerosol. *Communications Chemistry*, *1*(1), 75. https://doi.org/10.1038/s42004-018-0074-3

ESA. (2017). MERLIN (Methane Remote Sensing Lidar Mission) Minisatellite. Retrieved from https://directory.eoportal.org/web/eoportal/satellite-missions/content/-/article/merlin

European Environmental Agency. (2017, November 20). Contribution of the transport sector to total emissions of the main air pollutants [Data Visualization]. Retrieved November 18, 2018, from https://www.eea.europa.eu/data-and-maps/daviz/contribution-of-the-transport-sector-4

Faiz, A. (1996). *Air pollution from motor vehicles : standards and technologies for controlling emissions* (No. 16190) (p. 1). The World Bank. Retrieved from http://documents.worldbank.org/curated/en/606461468739185298/Air-pollution-from-motor-vehicles-standards-and-technologies-for-controlling-emissions

Gallagher, M. E., Down, A., Ackley, R. C., Zhao, K., Phillips, N., & Jackson, R. B. (2015). Natural Gas Pipeline Replacement Programs Reduce Methane Leaks and Improve Consumer Safety. *Environmental Science & Technology Letters*, *2*(10), 286–291. https://doi.org/10.1021/acs.estlett.5b00213

George, D. L., Poerner, N. W., Ridens, B. L., & Thorson, J. L. (2014). Compressed Natural Gas Vehicle Fuel Survey. Retrieved from https://trid.trb.org/view/1318095

Gilman, J. B., Lerner, B. M., Kuster, W. C., Goldan, P. D., Warneke, C., Veres, P. R., … Yokelson, R. J. (2015). Biomass burning emissions and potential air quality impacts of volatile organic compounds and other trace gases from fuels common in the US. *Atmospheric Chemistry and Physics*, *15*(24), 13915–13938. https://doi.org/10.5194/acp-15-13915-2015

Goldan, P. D., Kuster, W. C., Williams, E., Murphy, P. C., Fehsenfeld, F. C., & Meagher, J. (2004). Nonmethane hydrocarbon and oxy hydrocarbon measurements during the 2002 New England Air Quality Study. *Journal of Geophysical Research: Atmospheres*, *109*(D21). https://doi.org/10.1029/2003JD004455

Guisasola, A., Sharma, K. R., Keller, J., & Yuan, Z. (2009). Development of a model for assessing methane formation in rising main sewers. *Water Research*, *43*(11), 2874–2884. https://doi.org/10.1016/j.watres.2009.03.040

Halliday, H. S., Thompson, A. M., Wisthaler, A., Blake, D. R., Hornbrook, R. S., Mikoviny, T., … Hills, A. J. (2016). Atmospheric benzene observations from oil and gas production in the Denver-Julesburg Basin in July and August 2014. *Journal of Geophysical Research: Atmospheres*, *121*(18), 11,055-11,074. https://doi.org/10.1002/2016JD025327

Harley, R. A., Hannigan, M. P., & Cass, G. R. (1992). Respeciation of organic gas emissions and the detection of excess unburned gasoline in the atmosphere. *Environmental Science & Technology*, *26*(12), 2395–2408. https://doi.org/10.1021/es00036a010

Heath, G., Warner, E., Steinberg, D., & Brandt, A. (2015). Estimating U.S. Methane Emissions from the Natural Gas Supply Chain. Approaches, Uncertainties, Current Estimates, and Future Studies (No. NREL/TP-6A50-62820). National Renewable Energy Lab. (NREL), Golden, CO (United States). https://doi.org/10.2172/1226158

Heimann, I., Bright, V. B., McLeod, M. W., Mead, M. I., Popoola, O. A. M., Stewart, G. B., & Jones, R. L. (2015). Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. *Atmospheric Environment*, *113*(Supplement C), 10–19. https://doi.org/10.1016/j.atmosenv.2015.04.057

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., … Zhang, Q. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geosci. Model Dev.*, *11*(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Howard, C. J., Kumar, A., Malkina, I., Mitloehner, F., Green, P. G., Flocchini, R. G., & Kleeman, M. J. (2010). Reactive Organic Gas Emissions from Livestock Feed Contribute Significantly to Ozone Production in Central California. *Environmental Science & Technology*, *44*(7), 2309–2314. https://doi.org/10.1021/es902864u

Katzenstein, A. S., Doezema, L. A., Simpson, I. J., Blake, D. R., & Rowland, F. S. (2003). Extensive regional atmospheric hydrocarbon pollution in the southwestern United States. *Proceedings of the National Academy of Sciences*, *100*(21), 11975–11979. https://doi.org/10.1073/pnas.1635258100

Khordakova, D. (2017). *Diurnal behavior and temporal patterns of atmospheric peroxy radicals in a semi-urban area during summer 2016* (Master's). Institute of Environmental Physics, Bremen. Retrieved from http://www.iup.uni-bremen.de/PEP_master_thesis/thesis_2017/Khordakova_Dina_MScThesis.pdf

Klems, J. P., Pennington, M. R., Zordan, C. A., & Johnston, M. V. (2010). Ultrafine Particles Near a Roadway Intersection: Origin and Apportionment of Fast Changes in Concentration. *Environmental Science & Technology*, *44*(20), 7903–7907. https://doi.org/10.1021/es102009e

Korotcenkov, G., & Cho, B. K. (2011). Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey). *Sensors* *and Actuators B: Chemical*, *156*(2), 527–538. https://doi.org/10.1016/j.snb.2011.02.024

Lai, A. M., Carter, E., Shan, M., Ni, K., Clark, S., Ezzati, M., … Schauer, J. J. (2019). Chemical composition and source apportionment of ambient, household, and personal exposures to PM2.5 in communities using biomass stoves in rural China. *Science of The Total Environment*, *646*, 309–319. https://doi.org/10.1016/j.scitotenv.2018.07.322

Linquist, B., Groenigen, K. J. van, Adviento-Borbe, M. A., Pittelkow, C., & Kessel, C. van. (2012). An agronomic assessment of greenhouse gas emissions from major cereal crops. *Global Change Biology*, *18*(1), 194–209. https://doi.org/10.1111/j.1365-2486.2011.02502.x

Martenies, S. E., Milando, C. W., Williams, G. O., & Batterman, S. A. (2017). Disease and Health Inequalities Attributable to Air Pollutant Exposure in Detroit, Michigan. *International Journal of Environmental Research and Public Health*, *14*(10), 1243. https://doi.org/10.3390/ijerph14101243

McDonald, B. C., Gouw, J. A. de, Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., … Trainer, M. (2018). Volatile chemical products emerging as largest petrochemical source of urban organic emissions. *Science*, *359*(6377), 760–764. https://doi.org/10.1126/science.aaq0524

McDuffie, E. E., Edwards, P. M., Gilman, J. B., Lerner, B. M., Dubé, W. P., Trainer, M., … Brown, S. S. (2016). Influence of oil and gas emissions on summertime ozone in the Colorado Northern Front Range. *Journal of Geophysical Research: Atmospheres*, *121*(14), 8712–8729. https://doi.org/10.1002/2016JD025265

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., … Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, *70*(Supplement C), 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060

Mihelcic, J. R., Baillod, C. R., Crittenden, J. C., & Rogers, T. N. (1993). Estimation of VOC Emissions from Wastewater Facilities by Volatilization and Stripping. *Air & Waste*, *43*(1), 97–105. https://doi.org/10.1080/1073161X.1993.10467120

NASA. (2011, October 19). A Look Back at a Decade of Fires [Feature]. Retrieved December 10, 2018, from https://www.nasa.gov/mission_pages/fires/main/modis-10.html

Niedojadlo, A., Becker, K. H., Kurtenbach, R., & Wiesen, P. (2007). The contribution of traffic and solvent use to the total NMVOC emission in a German city derived from measurements and CMB modelling. *Atmospheric Environment*, *41*(33), 7108–7126. https://doi.org/10.1016/j.atmosenv.2007.04.056

Penza, M., Suriano, D., Pfister, V., Prato, M., & Cassano, G. (2017). Urban Air Quality Monitoring with Networked Low-Cost Sensor-Systems. *Proceedings*, *1*(4), 573. https://doi.org/10.3390/proceedings1040573

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., … Tans, P. (2012). Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study. *Journal of Geophysical Research: Atmospheres*, *117*(D4). https://doi.org/10.1029/2011JD016360

Pfister, G. G., Wiedinmyer, C., & Emmons, L. K. (2008). Impacts of the fall 2007 California wildfires on surface ozone: Integrating local observations with global model simulations. *Geophysical Research Letters*, *35*(19). https://doi.org/10.1029/2008GL034747

Phillips, N. G., Ackley, R., Crosson, E. R., Down, A., Hutyra, L. R., Brondfield, M., … Jackson, R. B. (2013). Mapping urban pipeline leaks: Methane leaks across Boston. *Environmental Pollution*, *173*, 1–4. https://doi.org/10.1016/j.envpol.2012.11.003

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., … Jones, R. L. (2018). Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmospheric Environment*, *194*, 58–70. https://doi.org/10.1016/j.atmosenv.2018.09.030

Ryerson, T. B., Aikin, K. C., Angevine, W. M., Atlas, E. L., Blake, D. R., Brock, C. A., … Watts, L. A. (2011). Atmospheric emissions from the Deepwater Horizon spill constrain air-water partitioning, hydrocarbon fate, and leak rate. *Geophysical Research Letters*, *38*(7). https://doi.org/10.1029/2011GL046726

Sabaliauskas, K., Jeong, C.-H., Yao, X., & Evans, G. J. (2014). The application of wavelet decomposition to quantify the local and regional sources of ultrafine particles in cities. *Atmospheric Environment*, *95*, 249–257. https://doi.org/10.1016/j.atmosenv.2014.05.035

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., … Zhu, Q. (2016). The global methane budget 2000–2012. *Earth System Science Data*, *8*(2), 697–751. https://doi.org/10.5194/essd-8-697-2016

Schultealbert, C., Baur, T., Schütze, A., & Sauerwald, T. (2018). Facile Quantification and Identification Techniques for Reducing Gases over a Wide Concentration Range Using a MOS Sensor in Temperature-Cycled Operation. *Sensors*, *18*(3), 744. https://doi.org/10.3390/s18030744

Sears, W. M., Colbow, K., & Consadori, F. (1989). General characteristics of thermally cycled tin oxide gas sensors. *Semiconductor Science and Technology*, *4*(5), 351. https://doi.org/10.1088/0268-1242/4/5/004

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollutaion to Climate Change* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Shamasunder, B., Collier-Oxandale, A., Blickley, J., Sadd, J., Chan, M., Navarro, S., … Wong, N. J. (2018). Community-Based Health and Exposure Study around Urban Oil Developments in South Los Angeles. *International Journal of Environmental Research and Public Health*, *15*(1), 138. https://doi.org/10.3390/ijerph15010138

Tan, Y., Lipsky, E. M., Saleh, R., Robinson, A. L., & Presto, A. A. (2014). Characterizing the Spatial Variation of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA. *Environmental Science & Technology*, *48*(24), 14186–14194. https://doi.org/10.1021/es5034074

US EIA. (2015, April 2). Top 100 U.S. Oil and Gas Fields. Retrieved December 12, 2018, from https://www.eia.gov/naturalgas/crudeoilreserves/top100/

US EPA. (2006). *Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final Report, 2006)* (Reports & Assessments No. EPA/600/R-05/004aF-cF). Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=149923

US EPA, OA. (2018). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016* (Reports and Assessments No. EPA 430-R-18-003) (p. 655). US EPA. Retrieved from https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2016

US EPA, OAR. (2015a, August 5). Managing Air Quality - Emissions Inventories [Overviews and Factsheets]. Retrieved December 20, 2018, from https://www.epa.gov/air-quality-management-process/managing-air-quality-emissions-inventories

US EPA, OAR. (2015b, December 3). Hazardous Air Pollutants: Sources and Exposure [Reports and Assessments]. Retrieved November 27, 2018, from https://www.epa.gov/haps/hazardous-air-pollutants-sources-and-exposure

US EPA, OAR. (2016a). *2014 National Emission Inventory (NEI) Report* (Policies and Guidance). Retrieved from https://www.epa.gov/air-emissions-inventories/2014-national-emission-inventory-nei-report

US EPA, OAR. (2016b, September 29). PAMS | Ambient Monitoring Technology Information Center | US EPA. Retrieved December 13, 2018, from https://www3.epa.gov/ttnamti1/pamsmain.html

US EPA, ORD. (2003, April 17). Benzene CASRN 71-43-2 | IRIS. Retrieved December 19, 2018, from https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?&substance_nmbr=276

USGCRP. (2018). Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. Washington, DC, USA: U.S. Global Change Research Program. Retrieved from https://nca2018.globalchange.gov/chapter/13/

Vergara, A., Llobet, E., Brezmes, J., Ivanov, P., Vilanova, X., Gracia, I., … Correig, X. (2005). Optimised temperature modulation of metal oxide micro-hotplate gas sensors through multilevel pseudo random sequences. *Sensors and Actuators B: Chemical*, *111*(Supplement C), 271–280. https://doi.org/10.1016/j.snb.2005.06.039

Wang, R., Moody, R. P., Koniecki, D., & Zhu, J. (2009). Low molecular weight cyclic volatile methylsiloxanes in cosmetic products sold in Canada: Implication for dermal exposure. *Environment International*, *35*(6), 900–904. https://doi.org/10.1016/j.envint.2009.03.009

Warneke, C., Geiger, F., Edwards, P. M., Dube, W., Pétron, G., Kofler, J., … Roberts, J. M. (2014). Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition. *Atmospheric Chemistry and Physics*, *14*(20), 10977–10988. https://doi.org/10.5194/acp-14-10977-2014

Westerholm, R., & Egebäck, K. E. (1994). Exhaust emissions from light- and heavy-duty vehicles: chemical composition, impact of exhaust after treatment, and fuel parameters. *Environmental Health Perspectives*, *102*(Suppl 4), 13–23.

Windischmann, H., & Mark, P. (1979). A Model for the Operation of a Thin-Film SnO x Conductance-Modulation Carbon Monoxide Sensor. *Journal of The Electrochemical Society*, *126*(4), 627–633. https://doi.org/10.1149/1.2129098

Wu, S., Mickley, L. J., Leibensperger, E. M., Jacob, D. J., Rind, D., & Streets, D. G. (2008). Effects of 2000–2050 global change on ozone air quality in the United States. *Journal of Geophysical Research: Atmospheres*, *113*(D6). https://doi.org/10.1029/2007JD008917

Yokelson, R. J., Burling, I. R., Gilman, J. B., Warneke, C., Stockwell, C. E., Gouw, J. de, … Weise, D. R. (2013). Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires. *Atmospheric Chemistry and Physics*, *13*(1), 89–116. https://doi.org/10.5194/acp-13-89-2013

Ziemann, P. J., & Atkinson, R. (2012). Kinetics, products, and mechanisms of secondary organic aerosol formation. *Chemical Society Reviews*, *41*(19), 6582–6605. https://doi.org/10.1039/C2CS35122F

CHAPTER 2

USING A LOW-COST SENSOR ARRAY AND MACHINE LEARNING TECHNIQUES TO DETECT COMPLEX POLLUTANT MIXTURES AND IDENTIFY LIKELY SOURCES

Abstract

An array of low-cost sensors was assembled and tested in a chamber environment wherein several pollutant mixtures were generated. The four classes of sources that were simulated were mobile emissions, biomass burning, natural gas emissions, and gasoline vapors. A two-step regression and classification method was developed and applied to the sensor data from this array. We first applied regression models to estimate the concentrations of several compounds and then classification models trained to use those estimates to identify the presence of each of those sources. The regression models that were used included forms of multiple linear regression, random forests, Gaussian process regression, and neural networks. The classification models that were trained included logistic regression, random forests, support vector machines, and neural networks. The best combination of models was determined by maximizing the F_1 score on ten-fold cross validation data. The highest F_1 score as calculated on testing data was 0.72 and was produced by the combination of a multiple linear regression model utilizing the full array of sensors and a random forest classification model.

1. Introduction

Understanding the causes of degraded air quality at a local scale is a challenging but important task. Making the task especially difficult is the complexity of possible sources of pollutants, each of which produce a wide variety of chemical species. These pollutant mixtures often include a component of volatile organic compounds (VOCs), which are important because of both direct toxicological health impacts as well as the impacts caused by secondary products like particulate matter (PM) and ozone. There is also no single instrument that can fully quantify the complex and dynamic changes of ambient air composition needed to identify local sources, and communities directly affected by air quality problems often do not have the resources to mount an extensive measurement campaign needed to identify likely sources. Quantifying the impact of and identifying these sources at a higher resolution than is currently possible would contribute to regulatory, scientific, and public health goals.

Regulators could benefit from this type of information by both understanding what types of sources to target in order to address specific problems and also by understanding what industries are likely to be affected by regulations targeting specific compounds. For example, a critical piece of legislation that governs toxic air pollutants is the list of Hazardous Air Pollutants (HAPs), which was produced and has been updated by the US EPA (US EPA, 2014). When determining whether a specific HAP is likely to affect a community, it is important to understand the nature of sources on a broad range of scales because some compounds only affect communities near to their source while others can have global effects. California's recent state bill, AB 617, is an example of the regulatory application of a low-cost measurement approach to understanding VOC sources. The bill is intended to address local air quality problems by both supporting community based monitoring of air toxics and using the data gathered by those monitors to address important sources of hazardous compounds (Garcia, 2017).

From a scientific perspective the location and nature of emission sources can provide important information for atmospheric chemistry models. These models are used to understand the production of important secondary products including forms of particulate matter (PM), ozone, and other compounds with health and environmental impacts. Currently, these models use some combination of emission factors, measurement campaigns, and back-propagation modeling to estimate emissions from important sources within urban areas. Although continuous measurements that are used as inputs to these models are often spaced on the order of tens to hundreds of kilometers apart, it has been shown that the concentration of chemically important compounds vary on the scale of single to tens of meters in urban environments (Cheadle et al., 2017; Tan, Lipsky, Saleh, Robinson, & Presto, 2014). Dense, continuous measurements of air quality within cities could provide insights into small-scale variations on a real-time basis that are difficult to capture using current techniques. For example, several studies have demonstrated the ability of dense networks or mobile platforms with low-cost sensors to quantify the variation of pollutants like NO_2, CO, O_3, and PM within urban locales (Mead et al., 2013; Penza, Suriano, Pfister, Prato, & Cassano, 2017; Popoola et al., 2018; Sadighi et al., 2018).

From a public health perspective, networks of low-cost sensors may provide the ability to understand real-life exposures for individuals to categories of compounds with suspected health impacts. The first of these studies using low-cost sensors have included some combination of small, portable measurement devices or dense networks and GPS locations of the individual. Once continuous and highly specific exposure data becomes available, it may be possible to couple that data with health outcomes to better understand the relation between exposures and outcomes at typical, ambient exposures (as opposed to workplace data, where exposure levels are typically much higher). Developing affordable methods to study air quality at local scales is important as it may help identify and address environmental justice problems that may contribute to higher disease burden in vulnerable communities (Martenies, Milando, Williams, & Batterman, 2017).

The work presented here provides a proof-of-concept for low-cost analytical tools to accomplish this goal on a more accessible scale. In our proposed methodology, analytical methods were applied to low-cost sensors in order to estimate the concentrations of important compounds and to predict the "class" of source that is emitting those compounds. Using this methodology, community members could have a "first pass" estimate of what they are being exposed to and what the likely source(s) of those compounds are. With these more human-readable outputs, community groups would be better able to make decisions such as when and where to take air samples, what pollutant species to focus on, or what part of their local government to reach out to for a follow-up measurement campaign with possible regulatory implications.

Previous studies have addressed components of the work entailed here but have not addressed the full scope of complex mixtures, ambient concentrations, and the prediction of realistic pollutant sources. For example, De Vito et al. have used arrays of low-cost sensors to detect and separate tertiary mixtures of compounds (Saverio De Vito et al., 2007; Saverio De Vito, Piga, Martinotto, & Di Francia, 2009), and others have applied similar approaches to develop "e-noses" that can identify specific mixtures of a few compounds (Akamatsu, Tsuruta, & Shin, 2017; Aleixandre et al., 2004; Capone, Siciliano, Bârsan, Weimar, & Vasanelli, 2001). The primary difference between these numerous studies and our own is that we have focused on identifying the likely source of the pollutants rather than attempting to identify individual compounds.

This study was also partially motivated by the problems understanding the outputs of low-cost sensors in the context of their increased noise and cross sensitivities relative to regulatory grade instruments (Clements et al., 2017). Our own field experience and recent papers from the low-cost air quality instrument community have consistently found difficulties understanding and interpreting the outputs of low-cost sensors, especially when communicating results to individuals who may not have a highly technical understanding of the instruments and/or atmospheric chemistry (Robinson, Kocman, Horvat, & Bartonova, 2018; Woodall et al., 2017). This study attempted to address some of these issues by providing an initial interpretation of the sensor outputs that can be produced in near real-time and are output in a way that a user could understand: the likelihood that a type of source is or is not affecting the measured air quality.

2. Materials and Methods

The methodology presented here was designed to accomplish two key goals. The first was to establish a system that is capable of detecting several compounds that are important air quality indicators. The second function of the system was to take those estimates and use them to identify likely source(s) of those pollutants. Specifically, this study aimed to develop a system that could identify pollutant sources that are often associated with VOC emissions. Note that the system here would not be more specific than a “class” of source like traffic, natural gas leaks, etc.

This can be understood to be similar to a typical measurement campaign intended to identify VOC sources. In these studies researchers would sample for several compounds of interest using research-grade instruments. Using those measurements, the researchers would then attempt to determine the likely sources based on their knowledge of typical emission factors and coemitted compounds.

To develop a preliminary low-cost source attribution tool, we assembled an array of diverse, low-cost sensors. We installed the array in an environmental test chamber and exposed it to simulated pollutant mixtures in a temperature- and humidity-controlled environment. We then adopted a two-step data analysis technique that would investigate the ability of the array to detect pollution and improve source attribution. The first step of this process was a regression model that estimated the concentration of pollutant species using the sensor array signal as input. We investigated several regression models that are discussed in the next section. The estimated concentrations were then used as the inputs to classification models that were trained to identify the source that was being simulated. Again, we investigated a diverse selection of classification models and compared the accuracy of source identification produced by each combination of regression and classification model.

2.1. Sensor Array Design

In order to detect a variety of gaseous compounds for inferring the presence of a local pollution source, it was necessary to develop an array of sensing elements that were sensitive to an equally varied mix of pollutants at relevant concentrations. This was accomplished by modifying the existing sensor platform, the Y-Pod (Ashley Collier-Oxandale, Coffey, Thorson, Johnston, & Hannigan, 2018; Piedrahita et al., 2014a). The Y-Pod is an Arduino-based open-source platform that is typically configured to include nondispersive infrared (NDIR), photoionization detector (PID), metal oxide (MOx), and electrochemical types of low-cost sensors for gaseous compounds (see Figure 1). This baseline configuration was added to through the use of two external boards made to house additional electrochemical and metal oxide sensors (see Figure 2). Two Y-Pods with this main board configuration and two additional boards were tested; some of the sensors varied between the two Y-Pods. The sensors included on each board are listed in Appendix A, and the data streams from both of these pods were combined and analyzed as if they were one large array.

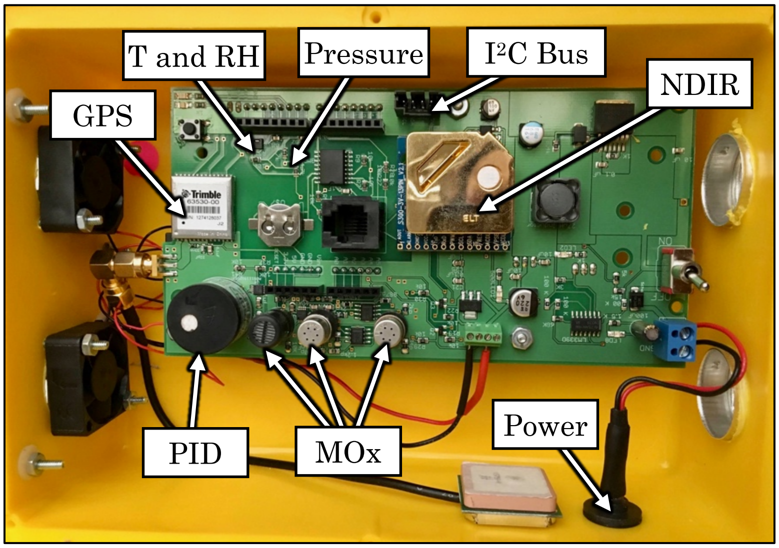
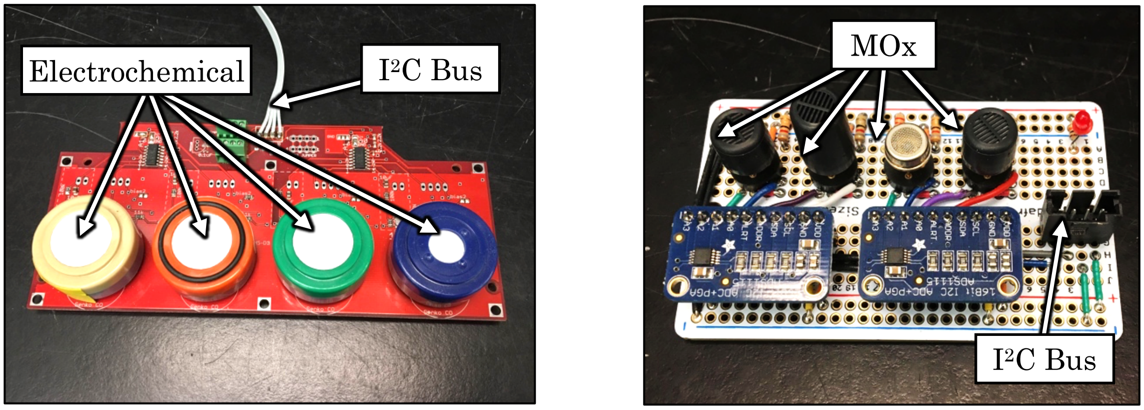


Figure 1. Image of a typical Y-Pod with important components labeled. The connector that is labeled "I^2C bus" was used both to communicate with ancillary sensor boards and to provide power to those boards. The boards were not installed in the ventilated box shown in this photo but were instead placed directly into the test chamber.



(a) (b)

Figure 2. Image of the additional boards that were used for electrochemical sensors (a) and metal oxide (MOx) sensors (b). The 4-wire connections labeled "I^2C Bus" contained both I^2C communications and electrical power for the boards.

It has been shown that changes in how sensors are operated and manufactured influence how those same sensors react to changes in gas concentrations and other environmental factors (G. Korotcenkov & Cho, 2011; Ghenadii Korotcenkov, 2013; Lee & Reedy, 1999; Schultealbert, Baur, Schütze, & Sauerwald, 2018; Tangirala, Gómez-Pozos, Rodríguez-Lugo, & Olvera, 2017; Yoo, 2011). With this in mind, sensors were selected from multiple manufacturers and with different target gases to maximize the variance of sensor signal dependence to different gases between individual sensors. One important aspect underlying the differences between sensors is the presence of non-specific cross-sensitivities that affect sensor signals when exposed to non-target gases (Alphasense Ltd., 2013, p. 109; Cross et al., 2017; Kim, Shusterman, Lieschke, Newman, & Cohen, 2017; Lewis et al., 2016a; Mead et al., 2013; Molino et al., 2012; Popoola et al., 2018; Spinelle, Gerboles, Kok, Persijn, & Sauerwald, 2017). When attempting to quantify a single compound, these interferences are treated as an issue that must be corrected for, but it is proposed here that a high dimensionality array of sensors could produce estimates of gases that are not the intended target of any single sensor. Some authors have investigated these cross-sensitivities and their usefulness when used in combination, but this field is relatively poorly understood (Hagler, Williams, Papapostolou, & Polidori, 2018).

Specific sensor models selected for use in this study (see Table 1) were determined after reviewing the literature to determine the sensitivity and usefulness of different models, as manufacturers rarely supply data from their sensors at concentrations that are relevant for ambient air monitoring. The variable name(s) associated with each of the selected sensors are listed in the table included in Appendix A.

Table 1. Details of the sensors that were selected for inclusion in the array.

Manufacturer	Model	Target Gas	Technology
Baseline Mocon	piD-TECH 0-20ppm	VOCs	Photoionization (PID)
ELT	S300	Carbon Dioxide (CO_2)	NDIR

Alphasense	H2S-BH	Hydrogen Sulfide (H_2S)	Electrochemical
	O3-B4	Ozone (O_3)	Electrochemical
	NO2-B1	Nitrogen Dioxide (NO_2)	Electrochemical
	CO-B4	Carbon Monoxide (CO)	Electrochemical
	NO-B4	Nitric Oxide (NO)	Electrochemical
Figaro	TGS 2602	“VOCs and odorous gases”	Metal Oxide
	TGS 2600	“Air Contaminants”	Metal Oxide
	TGS 2611	Methane (CH_4)	Metal Oxide
	TGS 4161	Carbon Dioxide (CO_2)	Metal Oxide
e2v	MiCS-5121WP	CO/VOC	Metal Oxide
	MiCS-2611	Ozone (O_3)	Metal Oxide
	MiCS-2710	Nitrogen Dioxide (NO_2)	Metal Oxide
	MiCS-5525	Carbon Monoxide (CO)	Metal Oxide

2.1.1. Metal Oxide (MOx) Sensors

Metal oxide-based low-cost sensors are extremely common in the literature due to their small size and especially low cost (on the order of $10 USD/ea.). They have been used to measure a range of species that are important for air quality measurement and include NO_2, O_3, CO, VOCs, and CH_4 (Ashley Collier-Oxandale et al., 2018; Fonollosa, Sheik, Huerta, & Marco, 2015; Molino et al., 2012; Schüler, Sauerwald, & Schütze, 2014; Spinelle, Gerboles, Villani, Aleixandre, & Bonavitacola, 2017; Vergara et al., 2005; Yoo, 2011). In the default configuration of the Y-Pod there are three “onboard” MOx sensors. Of these onboard sensors, two are designed for Figaro MOx sensors, and the heaters are connected to an op-amp circuit designed to keep the heater element at a constant resistance (i.e. temperature) by providing a varied electrical current. The third MOx pad is designed for an e2v sensor and drives the heater with a constant voltage. All three onboard MOx sensors’ sensing elements are measured using a voltage divider and analog to digital converter (ADC) chips. Additional MOx sensors were added to each Y-Pod board by installing them in voltage dividers with additional ADCs that communicated with the main board via the I^2C protocol (see Figure 2).

2.1.2. Electrochemical Sensors

Electrochemical sensors are another popular low-cost sensor technology and have been used widely for the study of pollutants at ambient concentrations (Borrego et al., 2018; Heimann et al., 2015; Kim et al., 2017; Lewis et al., 2016b; Mead et al., 2013; Molino et al., 2012; Popoola et al., 2018; Spinelle, Gerboles, Kok, et al., 2017; N. Zimmerman et al., 2017). For the purposes of this study, all electrochemical sensors were those manufactured by Alphasense and consisted of models designed to target NO, NO_2, CO, O_3, and H_2S with some sensor models duplicated. Alphasense electrochemical sensors were selected because of both the broad base of study that has been conducted with them in similar studies (Heimann et al., 2015; Kim et al., 2017; Mead et al., 2013; Molino et al., 2012; N. Zimmerman et al., 2017) and also because of the significant (claimed) variability in cross-sensitivities between the models selected for study here (Alphasense Ltd., 2013). As shown in Figure 2, the electrochemical sensors are all installed on boards containing a total of four sensors, each of which is in a potentiostatic circuit as recommended by their manufacturer (Alphasense Ltd., 2009) and the induced current is measured using an op amp circuit and ADC as also recommended by the manufacturer (Alphasense Ltd., 2009, p. 10).

2.1.3. Other Sensor Types

Nondispersive infrared (NDIR) and photoionization detection (PID) are two technologies that have been used in the literature for the detection of CO_2 and total VOCs (Lewis et al., 2016b; Popoola et al., 2018; Spinelle, Gerboles, Villani, et al., 2017). An NDIR sensor manufactured by ELT Sensor was included and was designed to be sensitive to CO_2. This sensor communicates estimated CO_2 concentration directly to the main Arduino board using the I^{2}C protocol. The included PID was manufactured by Baseline Mocon and is sensitive to total VOCs and outputs a proportional analog signal that varies from 0-2.5 volts as the concentration ranges from 0-20 ppm, respectively. The Y-Pod also includes commercial sensors for the measurement of ambient temperature, relative humidity, and barometric pressure.

2.2. Environmental Chamber Testing

For this study, it was determined that testing in a controlled environmental chamber would allow us to best characterize the array and to determine the feasibility of its usefulness when exposed to complex gaseous mixtures in the field. The test chamber used for this study was Teflon-coated aluminum with a glass lid with a total volume of approximately 1500 in^3. In order to reduce the total volume, two glass blocks were added to the chamber, reducing the unoccupied volume to 1320 in^3. A 40 mm, 6.24 CFM fan was also installed inside the chamber in order to promote mixing throughout.

Gas mixtures were varied in both total concentration and composition to attempt to capture the variation in emissions within different classes of pollution sources. Both temperature and humidity were adjusted independently to capture variability in sensor responses to both environmental parameters. The test plan was conducted intermittently over three months to attempt to capture temporal drift, a factor known to affect many types of low-cost sensors (Miskell et al., 2017; Romain & Nicolas, 2010). The full summary of concentration and other environmental parameter values produced during testing is available in Appendix B.

2.2.1. Simulating Pollutant Sources

We selected classes of sources to be simulated that are relevant to the communities in which these sensors could be deployed. The example community selected was one with which the Hannigan Group has performed other air quality research project and is located near downtown Los Angeles, California (Shamasunder et al., 2018). This community was selected because of its proximity to a variety of sources that may contribute to local air quality issues. These sources include oil and gas products and heavy traffic on major interstate highways, both of which can contribute to local elevations in concentrations of BTEX and other VOCs. Produced liquids from oil and gas extraction were simulated using gasoline headspace, which contains hydrocarbons could vaporize from produced oil and condensates. Additionally, natural gas leakage and low-NO_x combustion events (like biomass burning in wildfires) were selected as other sources relevant to the community.

Sources were simulated by injecting up to three component gases along with dilution air into the environmental chamber in a flow-through configuration. The dilution air for all testing was bottled "Ultra Zero" air, which is a synthetic blend of oxygen and nitrogen and contains less than 0.1 ppm of total hydrocarbons (Airgas, 2018). Flow control was accomplished using mass flow controllers (MFCs) that ranged from 0-20,000 SCCM for the dilution air and 0-20 SCCM for the component gases. Each MFC was calibrated in situ using bubble flow meters as reference flow measurements, and a linear calibration was applied to each MFC prior to calculating in-chamber concentrations. Reference gas concentrations were calculated as a flow-weighted average as shown in Equation 1 where $F_{i,t}$ is the concentration of the compound at the inlet of the chamber, $F_{i,b}$ is the concentration of the pollutant in the supply bottle, $Q_{i,b}$ is the flow rate of the gas from that bottle, and Q_m is the flow rate as reported by each MFC, including the dilution air.

$$F_{i,t} = \frac{\sum_0^j F_{i,b} * Q_{i,b}}{\sum_0^n Q_m} \tag{1}$$

Our test chamber did not include instrumentation to measure concentrations within the test chamber, so we assumed that the concentration of gas flowing into the chamber was equal to the concentration within the chamber. In order to make this a reasonable assumption, we determined the time the chamber needed until the concentrations had stabilized within the chamber after any inlet changes. This stabilization time period was calculated as four times the time constant for the chamber, a value that was determined to be roughly 600 seconds at 8.5 L/min, the nominal flow rate used for testing.

Table 2. Simulated source classes and the species used to simulate them for the purpose of this chamber testing. Note that limited dilution ratios did not allow for the inclusion of VOCs in the combustion sources.

Source	Component Gases
Biomass Burning	CO, CO_2
Mobile Sources	CO, CO_2, NO_2
Gasoline/Oil and Gas Condensates	Gasoline Vapor
Natural Gas Leaks	CH_4, C_2H_6, C_3H_8

The component gases are listed in Table 2 for each simulated source and timeseries of their concentrations are illustrated in Figure 3. For the testing of each source, the total concentration of the mixture and the relative concentrations of each component gas were varied to attempt to better simulate variation within source classes while still remaining representative of the source. For example, the level of nonmethane hydrocarbons (NMHCs) within the simulated natural gas was varied from 0%, representing an extremely dry gas to just over 20%, representing a rich gas that would be more representative of midstream or associated natural gas. This variation was also intended to try to remove some of the correlations between component gases to allow better understanding of what sensors were specific to which gas. The dilution ratio and ratios of component gases were limited by the flow ranges of the MFCs and the concentrations of the gas standards used to supply the component gases. For a full matrix of conditions for each test point, please refer to Appendix B.

The two simulated combustion sources did not include VOC species despite the associated VOC production motivating their inclusion in this testing. This decision was made for two reasons. First, the typical emission rates of primary compounds like CO, CO_2, and NO_2 from highway traffic are an order of magnitude higher than the emission rates for total gaseous VOCs on a mass basis, and even higher on a molar basis (Westerholm & Egebäck, 1994). Secondarily, the NMHC emissions from combustion are dominated by highly reactive oxygenated volatile organic compounds (OVOCs) as well as semi-volatile organic compounds (SVOCs) (Akagi et al., 2011), both of which are impractical to obtain as reference gas cylinders because of their propensity to react and/or condense. Because of these factors, it was decided to attempt to detect the primary pollutants (CO, CO_2, and NO_2) as indicators for the two types of combustion sources. NO_x was not included with "low temperature combustion" both because NO_x emissions are dominated by diesel combustion (US EPA, 2016) and because it has been shown to quickly react with other emitted species into PAN and particulates (Yokelson et al., 2013). If a significant influence of mobile sources or biomass burning was detected, one could use typical emission rates to estimate VOC exposures and/or could bring a more precise and targeted instrument into the community to make regulatory or other important decisions.

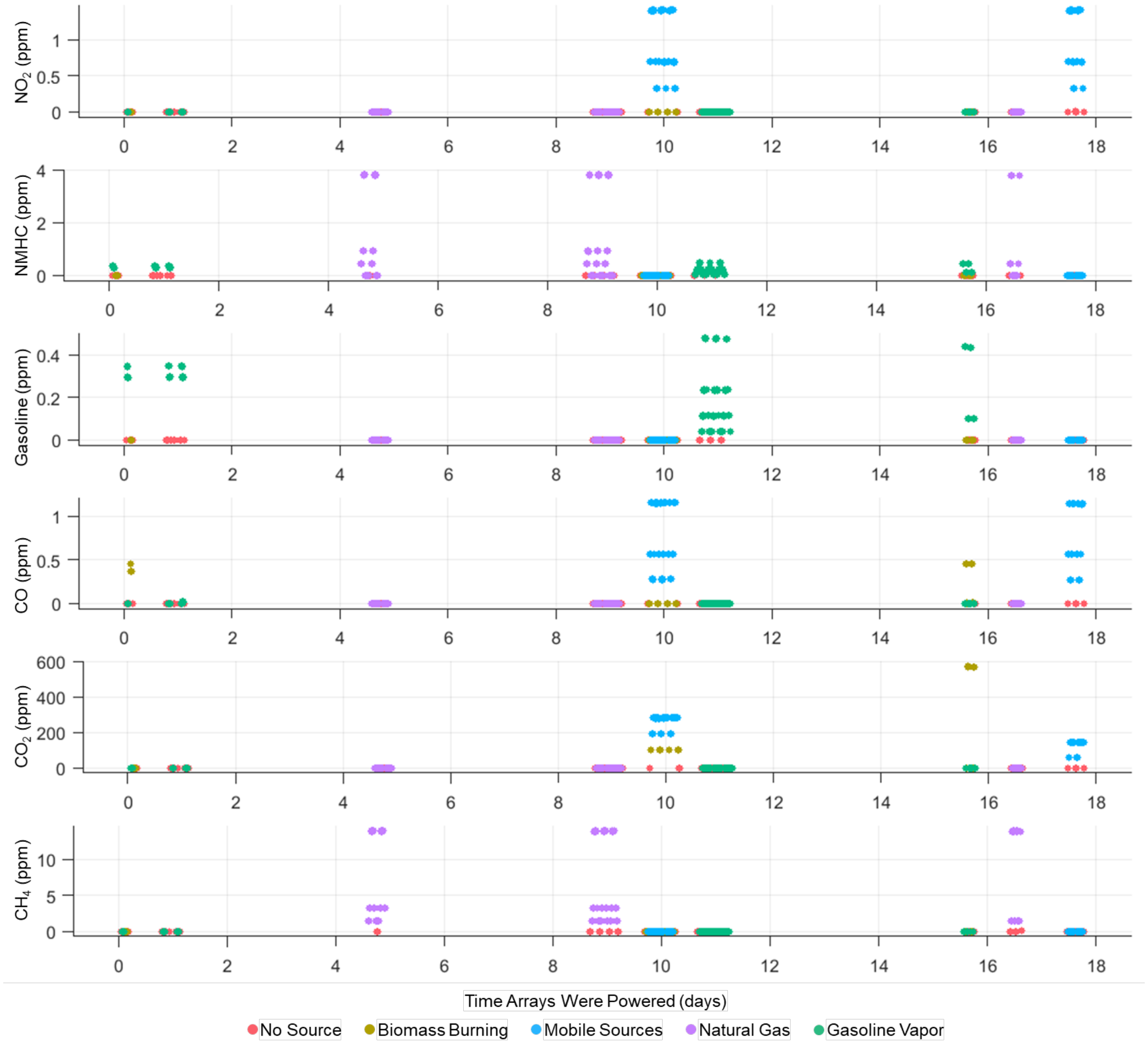


Figure 3. Timeseries of reference concentrations for each component gas, colored by the simulated pollutant source. X-axis indicates elapsed time in days that the sensor arrays were powered. Gaps in the plots indicate times that they were powered but no test was being performed. During these times, the arrays were left in the chamber at room temperature and exposed to ambient indoor air.

2.2.2. Simulating Other Environmental Parameters

In addition to simulating different mixtures of gases, temperature and humidity were also varied to attempt to capture environmental effects that are known to affect low-cost sensing devices (Cheadle et al., 2017; Suárez, Arroyo, Lozano, Herrero, & Padilla, 2018; Tsujita, Yoshino, Ishida, & Moriizumi, 2005). Temperature control was accomplished using an AC resistive heater, and humidity control was accomplished by bubbling the dry dilution air through a water bath. Typical values ranged from 25-45°C and 40-80%, respectively. Both temperature and humidity were recorded inside the chamber throughout testing, and a 12 VDC fan was installed inside the chamber to promote mixing.

2.3. Computational Methods

After subjecting the sensor arrays to a range of mixtures and environmental conditions, it was possible to develop training algorithms to quantify the component gases and predict the presence of each simulated source. To begin this process, the data from both the sensor arrays and the chamber controls were collated and prepared for analysis. Because of the high dimensionality and nonlinearity of the sensor dataset, a two-step approach was adopted and is illustrated in Figure 4. In this approach, a regression model is first fitted to predict the concentration data for each gaseous compound or class of compounds, e.g. carbon monoxide or NMHC. The application of this regression model created a pre-processing step that reduced the dimensionality of the dataset in a way that created new features that were useful on their own. The estimated concentrations of each compound were then used as features for a classification model that attempted to predict the class of source that was being simulated, e.g. "Natural Gas" or "Mobile Sources". The classification models for each source were trained to allow for the prediction of multiple simultaneous sources, although multiple source simulations were not tested in this capacity due to limitations on the number of MFCs in our test chamber.

Figure 4. Information flow diagram illustrating how predictions of the sources of detected pollutants (in this case, natural gas) are generated indirectly from raw sensor data.

2.3.1. Initial Data Preprocessing

Both datasets were down-sampled to their 1-minute median values to remove noise and transient changes in conditions that could not be assumed to be uniform throughout the chamber. During preprocessing, other data cleaning steps included the removal of 60 minutes after the pods were turned on (warm up times). The temperature and relatively humidity values were also converted to Kelvin and absolute humidity, respectively., the data from the stabilization period described earlier were removed. This ensured that enough time was allowed for the chamber to reach its steady state such that it was appropriate to assume that the inlet mixing ratio was representative of the actual mixing ratio within the chamber.

2.3.2. Validating Models

Because environmental data is highly autocorrelated, it was important to select holdout data for validation that would better represent unseen data than simple random sampling. For this study, holdout data was selected using k-fold validation with 10 folds as illustrated in Figure 5. Using this method, the full dataset was sorted by date and time and then divided into ten sequential blocks, or folds, of data. For all folds, each model was trained on the nine other folds of data and tested on the fold of interest. If model parameters were optimized for a given model, a subset of the training data was used for validation during that optimization process. The k-fold validation process produced ten estimates for each concentration and source at each timestep – one produced by each model trained on a subset of the data. This provided some indication of the sensitivity of the models to the set of data on which it was trained and therefore provides an indication of how well the model could generalize to new data relative to the other models.

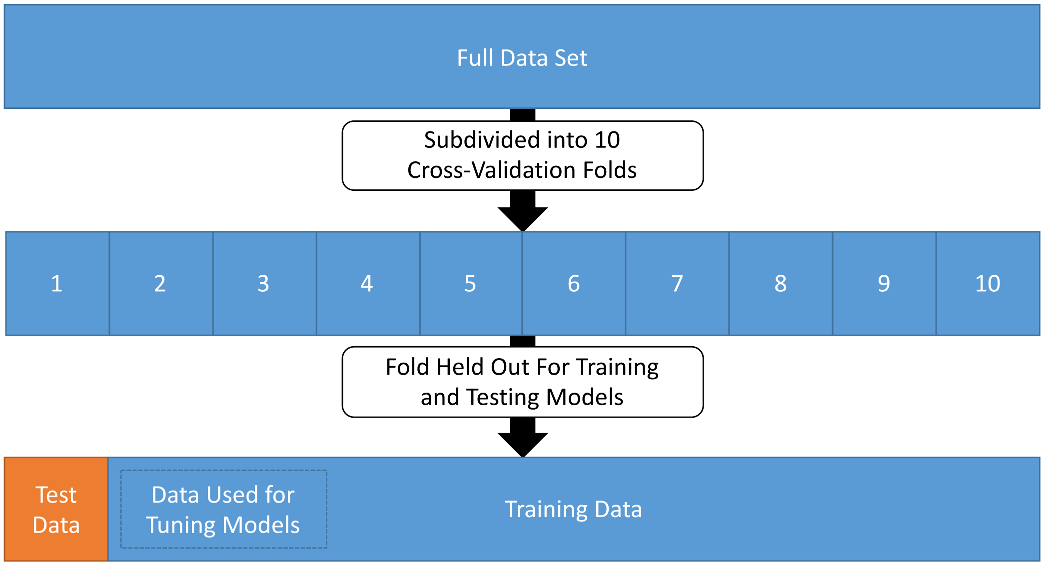


Figure 5. Graphic illustration of the 10-fold cross-validation method used here illustrating the use of the first fold for testing. Each fold contained 10% of the data, selected as consecutive blocks of time from start to end. Data used for validation when tuning hyperparameters were a subset of the training data (models were never exposed to the test data during training). For each cross-validation fold, classification models were tested on data that was also in the "test data" set during regression model training.

2.3.3. Estimating Concentrations with Regression Models

The first step in the analysis process involved regressing several "key" compounds that were selected as especially important for both detecting and differentiating the emissions from different classes of sources. Several regression techniques that have been applied in the literature were explored here and assessed first on their ability to accurately model the concentrations of several different gases. These regression techniques included multiple linear regression (MLR), ridge regression (RR), random forests (RF), gaussian process regression (GPR), and neural networks (NN). Separate models were trained for each compound to avoid "learning" the artificial correlations that were present in this study but would not fully represent the diversity of mixtures that could be expected in a field deployment.

Some models include terms referred to here as "hyperparameters". Hyperparameters govern how the model operates or is trained and are set before beginning to train the model. For example, the number of layers and nodes in a neural network would be considered a hyperparameter. Another example is the regularization strength (k) in ridge regression, which modifies the loss function used while training the model. Tuning hyperparameters is one way to improve the performance of models and can be used to determine appropriate values for model parameters that might normally be selected arbitrarily or based on previous experience. Wherever possible, these hyperparameters were optimized to maximize the estimated model performance on future data. This was done by holding out a subset of the training data and optimizing the hyperparameters based on the model performance when validated on that subset of the training data. This data selection is illustrated in Figure 5. Once the hyperparameters were set, the regression models were also assessed based on how well the classification models in the next step were able to use the estimated concentrations to make a prediction of the "source" that was being simulated.

Multiple linear regression is one of the most popular forms of regression used to convert sensor signal values into calibrated concentration estimates and have been used in a range of applications with many low-cost sensor technologies (Bigi, Mueller, Grange, Ghermandi, & Hueglin, 2018; Casey, Collier-Oxandale, & Hannigan, 2019; Cheadle et al., 2017; S. De Vito et al., 2018; Kim et al., 2017; Spinelle, Gerboles, Villani, et al., 2017). Because of popularity and the relatively low computational costs, several forms of multiple linear regression were investigated. The first form, referred to here as "FullLM" was a multiple linear regression model that simply used every sensor input from the array. This was considered as a baseline model as it included almost no previous knowledge except for the design of the array itself. Next was a model referred to from now on as "SelectLM". These models only included only data from sensors that were known to react to the current target gas, as determined by a combination of field experience and manufacturer recommendations. For each sensor included in a model, an interaction term between the sensor response and the measured temperature was included. Humidity was also included in each model, but no interaction term was added.

Stepwise linear regression ("StepLM") is an interesting but slow to train method of determining important predictors for use in a linear regression model. In this methodology, a base model (typically a simple intercept term or a full interaction model) is fit to the data. In the case of this study, the initial model was a simple intercept because a full interaction model was prohibitively large and slow. After fitting the original model, new models are trained with the addition or removal of one of each possible term. If the addition of a new term improves upon the previous model above some threshold, or if the removal does not reduce performance by a similar threshold, the term is added or removed. This new model is used as the new "base" model and the process is iterated until no terms can be added to or removed from the model within those constraints. In this study, the metric for improvement was the R^2 value, and the threshold was a value of 0.075, which was chosen by experimenting with different performance metrics and exploring the performance of the generated models.

Ridge regression ("RidgeLM") is a form of multiple linear regression, but in this study the two are differentiated by how the terms of the multiple linear regression were determined. During the generation of ridge regression models, all sensor values were included, and additionally, the interaction between each sensor and both temperature and humidity were included. This created a high dimensionality dataset, on which ridge regression was applied; a method that includes a term to penalize overfitting. Increasing this term affects the loss function, so that the final model assigns low weights to sensor signals that are not generally useful. This value of k was determined during initial investigations and then was kept constant across different compounds and cross-validation sets.

Outside of these linear regression models, several nonlinear models were trained. These were random forest regression ("RandFor"), gaussian process regression ("GuassProc"), and neural networks ("NeurNet"). For random forest regression, a large ensemble of decision trees is trained to output discrete values. These trees are each trained on different bootstrap aggregated ("bagged") subsets of the original data that are selected randomly with replacement. When making predictions, the outputs of each of these trees are averaged in order to produce an output that can approximate a smooth function. Because the individual trees only learn to produce values that they have seen before, the extension of random forests outside of their training space may be limited, although Zimmerman and colleagues showed that they were able to produce reasonable results within some parameter space (Naomi Zimmerman et al., 2018). The hyperparameters that were optimized for random forests were the minimum number of points at each leaf node, the maximum number of splits for each tree, and the number of variables to select from at each edge. These were optimized using the loss function out-of-bag error, which is the error on data that were not selected during the "bagging" of data during initial training.

Gaussian process regression, which is sometimes referred to as kriging, is a probabilistic method that uses training data and some assumptions about the distribution of the variable to make predictions on new data. Because this method is nonparametric, the ability to extrapolate to new data is somewhat limited, however, it is popular in the environmental modeling community and De Vito recently showed good success applying them to real atmospheric data (S. De Vito et al., 2018). A squared exponential kernel was used after some initial investigation showed little dependence of the results on this selection. The hyperparameters that were programmatically optimized for this model were the kernel parameters and were optimized by minimizing the objective function: log(1 + cross-validation loss).

Finally, the last regression model explored here was a neural network, a technique that has been used with low cost sensor for some time but is seeing a resurgence as improved training methods and computational power have improved their applicability (Casey et al., 2019; S. De Vito et al., 2018; Szczurek, Szecówka, & Licznerski, 1999). These models produce results by combining a set of "neurons" into a larger network. Each neuron applies a set of weights to each input and use a transfer function to translate the sum of those inputs into an output for the neuron. The first layer of neurons uses the raw sensor values as inputs, and subsequent layers use the outputs of the first layer as inputs. These two layers are often referred to as hidden layers, the last of which provides the input to the output layer that translates these outputs into a single predicted value. The hidden layers and number of nodes in each hidden layer of a neural network model are tunable hyperparameters and were optimized for the best mean squared error (MSE) on a subset of the training data that was held out for testing. The number of nodes in each hidden layer was varied between 1-40 for the first layer and from 0-40 for the second layer. When the number of nodes in the second layer was specified as "0", the second hidden layer was simply omitted.

2.3.4. Predicting Source Classes with Classification Models

After generating estimates of "key" compounds using each of the above regression approaches, classification algorithms were trained to identify the class of "source" that was being simulated, using the estimated concentrations at each timestep as features. The data was divided into the same sets of calibration and validation sets to ensure that the final validation results were left out of model training for both regression and classification. The classification techniques applied here have been seen in the literature, although typically with the goal of identifying individual compounds within simple mixtures (Saverio De Vito et al., 2007; Marco & Gutierrez-Galvez, 2012; Tomchenko, Harmer, Marquis, & Allen, 2003; Vembu, Vergara, Muezzinoglu, & Huerta, 2012). Those algorithms selected here are logistic regression, support vector machines (SVM) with both a linear and Gaussian kernel, random forest classifiers, and neural networks. The models vary significantly in their ability to separate different classes and were selected because of that diversity. For all of the methods presented here, the classification model outputs were values ranging from 0 to 1, where 0 indicates high confidence in the absence of a source, and 1 indicates high confidence in the presence of a source. When comparing the results to the actual simulated source, a value greater than or equal to 0.5 indicated a prediction that the source was present and a value less than 0.5 that it was absent.

The first type of classifier, the logistic regression is the simplest and most linear of the classifiers. In the results below, this model was referred to as "Logistic_class" and is a generalized linear model with binomial distribution. An independent classifier was created for each of the simulated sources, and each model was trained to indicate the presence or absence of that source. Much like the StepLM function described earlier, the terms for this model were selected by stepwise regression, with the difference being that the terms here were gas concentrations rather than sensor signal values. Because the logit link function is used to map the output of a linear function to a value from 0 to 1, the output of a logistic regression is often interpreted as the probability or confidence that the value is in the positive or negative class. In this case, that would be the likelihood that a certain source is or is not affecting the measured air quality. One benefit of logistic regression is that it is interpretable and computationally inexpensive relative to other, more complex and nonlinear methods.

The second classification model investigated was support vector machines (SVMs), which were trained to indicate the presence of a source. The general goal of SVMs is to create a line or plane that has the largest margin between separated classes, with some allowance for outliers and noise. In two-dimensional space, this can be visualized as creating a line that separates two classes and has the widest empty space on either side. The points closest to the line are referred to as support vectors and give the classifier its name. Because the goal of an SVM is to create a line that separates two classes, it may be considered as a linear classifier, although kernel functions are often used to map this linear function to a nonlinear space. The new features created by kernel methods are generally understood as measures of similarity between each instance. Two kernels were studied here, the first being a linear kernel that does not transform the input variables, and the second being the gaussian kernel. These are referred to in the results as "SVMlin_class" and "SVMgaus_class" respectively, simply because of the name of the functions written to implement them. Both kernels were implemented to predict the presence or absence of each source independently, just as described for logistic regression above. For the SVM with a gaussian kernel, the hyperparameters controlling the kernel scale and box constraint were optimized automatically to reduce cross-validated misclassification errors. These two factors affect how "smoothed" the kernel is and how heavily the loss function is penalized for errors, respectively.

Next, a random forest classification model was trained to identify the presence of each source, referred to in the results as "RandFor_class". Although random forests can quite easily be applied to multiclass regression problems, models were trained to predict the binary presence of each source separately, so that in sum the models could predict the presence of multiple sources at the same time without creating many classes representing all possible combinations of sources. The hyperparameters that were optimized for this model were the same as for the random forest regression models: minimum leaf size and maximum number of splits. These random forests represent a collection of 500 decision trees that are each trained on different subsets of the training data. In this case, those subsets were selected via bootstrap aggregation (bagging), wherein data is randomly sampled with replacement from the original training dataset. Each of these trees produce a prediction of the presence or lack thereof for the source that they were trained on. The "score" that was output to indicate the confidence that a source was present represents the fraction of trees within each forest that predicted that a source was present.

The last classification approach that was explored was a pattern recognition neural network, referred to in the results as "NeurNet_class". This neural network had one output node for each source and was therefore able to predict the presence of multiple sources independently and simultaneously. The hyperparameters that were optimized for this model were the number of layers (1 or 2), the number of nodes in each layer, and the transfer function that translates the input to outputs. In this case a neural net with 2 layers of 5 nodes using the "softmax" transfer function was selected.

Other classification methods that were not included here but are present in the literature include k-nearest neighbor classifiers (Schüler, Fricke, Sauerwald, & Schütze, 2015; Suárez et al., 2018), linear discriminant analysis (Bastuck, Bur, Lloyd Spetz, Andersson, & Schütze, 2014; Romain & Nicolas, 2010; Schüler et al., 2014), and other, more specialized techniques targeted at specific sensor technologies (Lee & Reedy, 1999; Schüler et al., 2015). These were not included due largely to their similarities to other techniques applied here.

3. Results

3.1. Regression Results

After training each model, the accuracy was measured as the root mean squared error (RMSE). Other analysis metrics like, R^2, that use the variance of the reference values were difficult to interpret because some folds of validation data consisted entirely of concentration values that were exactly zero. In other words, some validation folds consisted of time periods when a compound was not being added to the chamber. The RMSE values plotted in Figure 6 illustrate that the models perform acceptably but are generally not as accurate as has been achieved in the literature (Borrego et al., 2018; Casey et al., 2019; Mead et al., 2013; Spinelle, Gerboles, Kok, et al., 2017). They also show that they are susceptible to overfitting to various degrees as shown by the differences between the results on training data (dark colored boxes) and testing data (light boxes). The relatively poor performance that is reflected in the RMSE values is also reflected when plotting the estimated concentrations versus reference values. This is shown in in Appendix C.

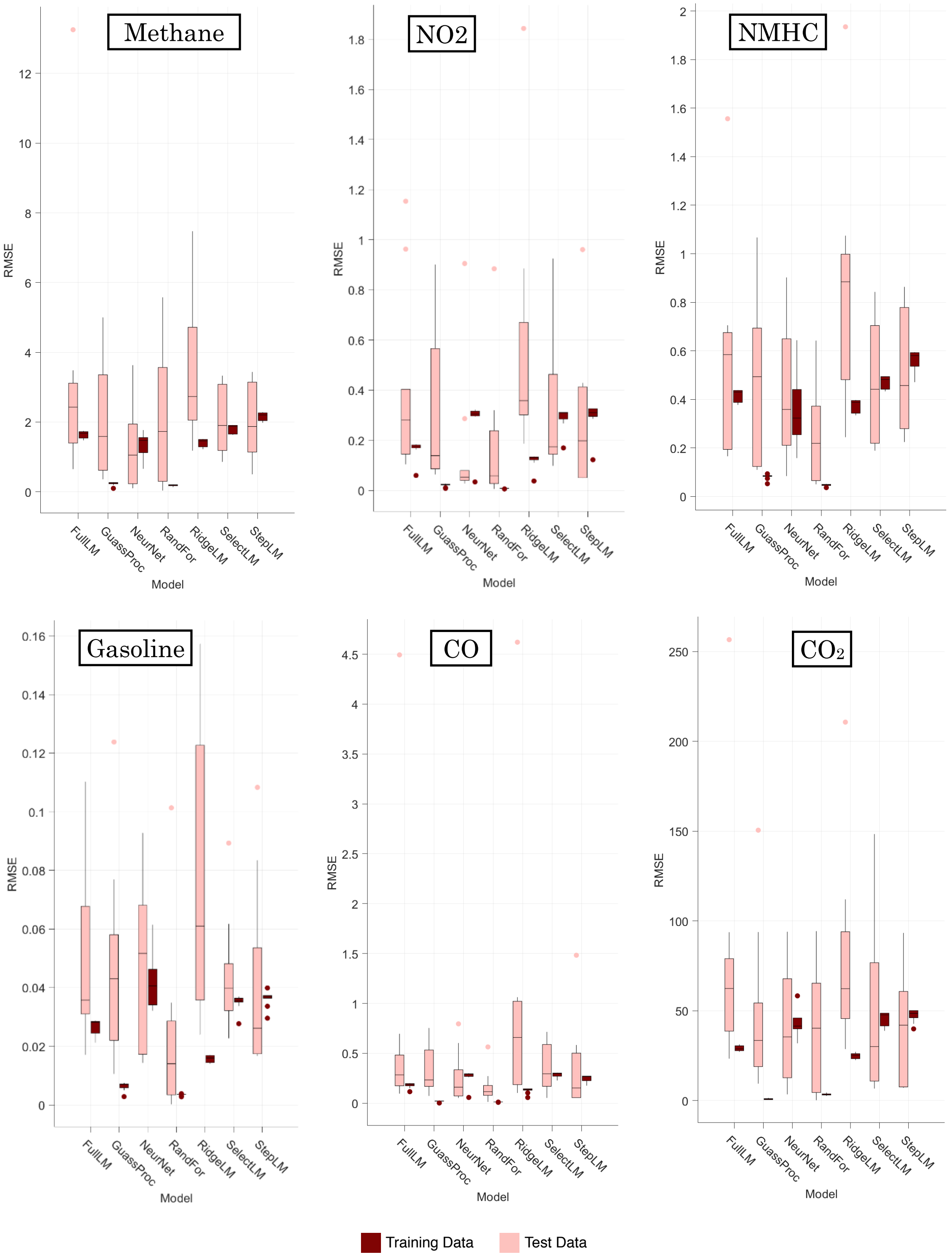


Figure 6. Boxplots of RMSE values (ppm) from each fold using each model and species. Bars indicate the 25th and 75th percentile values, whiskers extend to values within 1.5 times that range from the median, and dots represent points outside of that range as calculated using the GRAMM package for MATLAB (Morel, 2018).

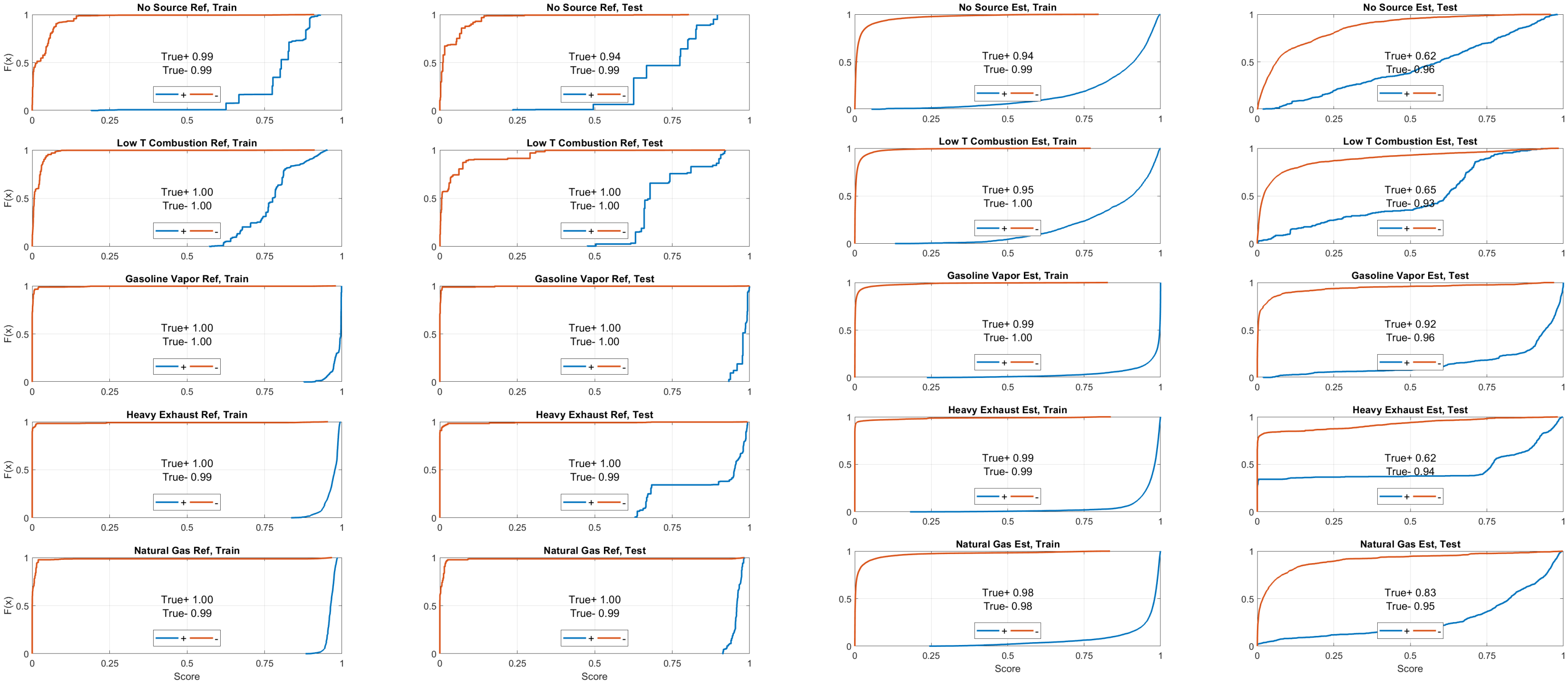
Interestingly, the variation between models when compared by their RMSE is not as great as one might expect. It also appears that the regularization as applied to the ridge regression did not totally address the problem of overtraining, as evidenced by the training RMSE values consistently underestimating the testing errors. This indication of overtraining could indicate that the regularization didn't balance out the increased variance caused by the addition of temperature and humidity interaction features introduced only for ridge regression. Neural nets and random forest regression typically performed best in our study, possibly explained by their ability to encode complexities like interactions, changes in sensitivity over time, and nonlinearities in signal response.

The relatively poor performance of the models, as compared to the literature, may be partially explained by design of the experiment. That is, this chamber study was designed to intentionally introduce a wide range of conditions containing known or suspected confounding gases at significantly elevated levels. This wide range of sources and concentrations would be unlikely to occur in a single location and it was therefore more challenging to accurately predict concentrations than in a typical environment. Reference values are also calculated based on MFC flow rates, which introduces the assumption that the chamber is fully mixed and has reached steady state concentrations equal to that at the inlet. Both of these assumptions necessarily introduce additional uncertainties. Finally, the nature of chamber studies produces relatively discrete values for each reference instead of the continuous values that would be experienced in an ambient environment. This forces models to interpolate across a relatively wide range of conditions that are not seen in the training data; likely an important factor that may reduce the accuracy of the models.

3.2. Classification Results

At the end of the classification step, a set of values from 0 to 1 was output by each of the classification models for each simulated source. These values indicated the model's prediction that a source was or was not present at each time step. In a field application, these values would be the indication of the likelihood that a source was affecting the measured air quality.

The results of each combination of classification and regression model were evaluated using empirical cumulative probability functions (CDFs) of the model outputs when a source was and was not present. An example of one of those plots is shown in Figure 7 for the best performing set of models ("FullLM" and "RandFor_class"). By plotting a CDF of the output of the model, it was possible to see how often a model correctly predicted the presence of a source (score ≥ 0.5) or absence of a source (score < 0.5) when the source was present (blue line) and was not present (red line). The magnitude of the score also indicated how confident the model was in its prediction, so values closer to 0 or 1 would be more useful to an end user because it would give a stronger indication that the source was actually absent or present. For example, if the classifier output a value close to 1 for traffic emissions, the user could be more confident that elevated concentrations were due to an influence of traffic pollution. In these plots, an ideal and omniscient model would roughly trace a box in the plots below, where it output a score close to 0 for all points where a source was not being simulated (red), and close to 1 for all points when a source was being simulated (blue). The CDF was plotted for classification models trained on both the reference values and the estimated values output by regression models. This allowed us to determine whether a model would or would not be able to separate the sources when given "perfect" concentration values. If not, then the model was unlikely to be useful when trained on less accurate estimates produced by regression models. If the model performance was adequate when trained and validated on reference data, then it was trained using the outputs of each regression model to see which it was most compatible with.



(a)

(b)

Figure 7. Example set of CDFs showing classification model performance on both calibration (train) and validation (test) data. The red lines trace the CDF for test cases where the source was not present (-), and blue lines illustrate the results when the source was present (+). (a) shows performance using the reference gas concentrations as features, and (b) shows performance for a combination of regression and classification models (in this case, FullLM and RandFor_class).

Another analysis tool that was used to select models were the confusion matrices that are shown in Figure 8 for each combination of regression and classification models. The confusion matrices plotted here illustrate the fraction of times each source was identified as present when each of the sources was being simulated. Confusion matrices are useful because they help us to understand what types of mistakes the model is making. For example, the top row of each confusion matrix shows the fraction of times that a model predicted that a source was present when no source was being actively simulated. A confusion matrix for an ideal model would be a matrix with the value of 1 along the diagonal and 0 elsewhere, indicating that it predicted the correct source every time. As mentioned before, the threshold was set as 0.5 for each model output, and each model was trained independently, so it was possible that multiple sources were predicted for a single timestep. This functionality was intentional because it would allow for the identification of multiple simultaneous sources, a situation that is likely to occur in reality.

Several of the models had problems with confusion between low and high temperature combustion that should have been easily differentiated by the presence of NO_x. After reviewing the data, it appears that the NO_2 electrochemical sensor failed approximately 2/3 of the way through the test matrix, likely hampering the ability of regression models to learn its importance in predicting NO_2 concentrations. Interestingly, the regression model that used all of the sensors was the one that produced the best inputs to the classifier models. This may be because it did not discard other sensors that would have been possibly less useful for the regression of NO_2 but that also did not fail near the end of the test.

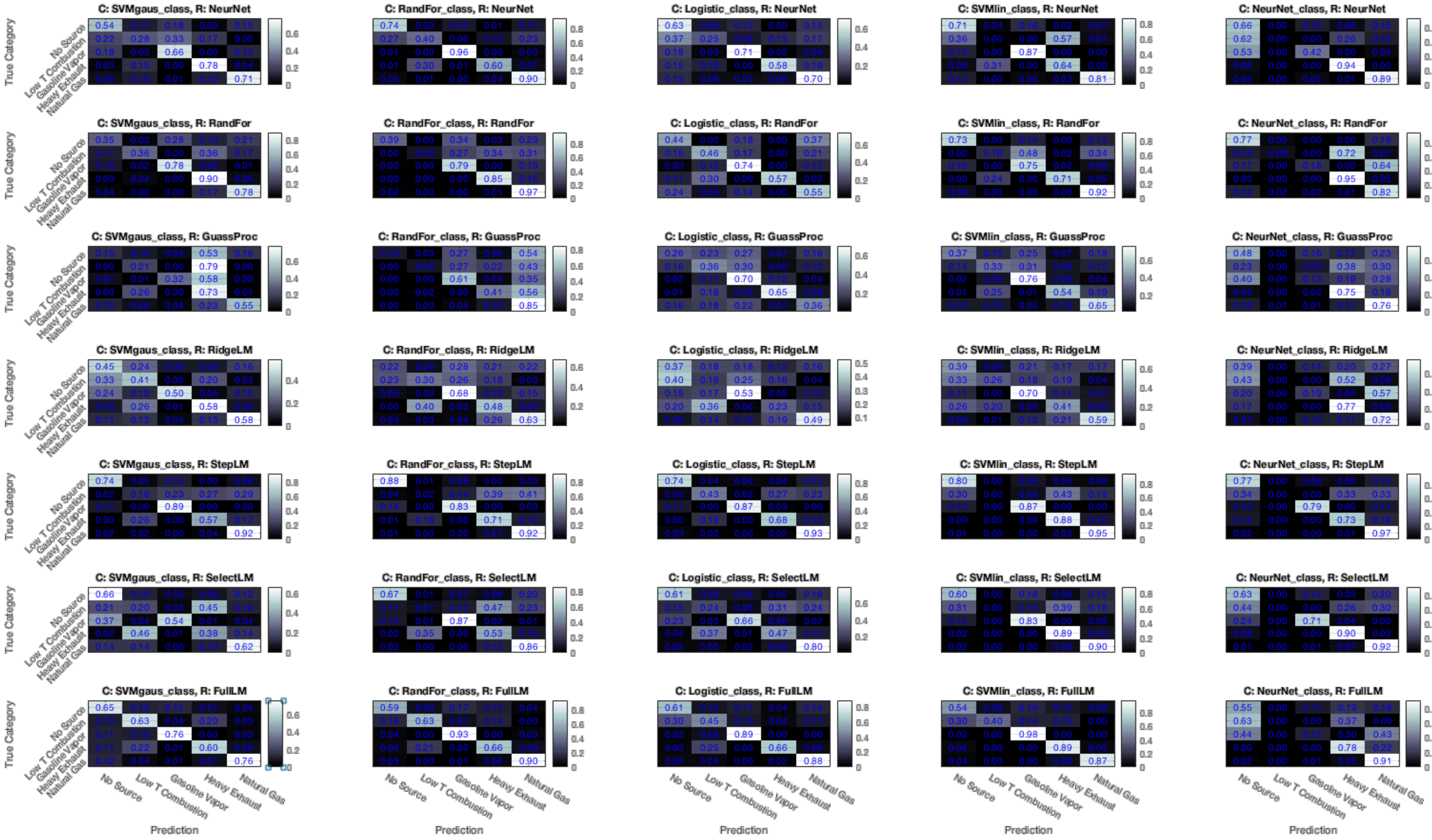


Figure 8. Confusion matrices showing the classification success on validation data with the F_1 score noted for each (see equations 1-4). Each chart shows the results of a combination of classification (columns) and regression (rows) models. Within each confusion matrix, the reference class is shown on the "y" axis and predicted class is indicated on the "x" axis. Lighter colors indicate higher values, ranging from 0 to 1, which correspond to the fraction of times that a source was predicted when a source was being simulated.

4. Discussion

4.1. Best Combination of Models

The best combination of classification and regression model was determined by comparing F_1 score as calculated using equations (1-4). The F_1 score is a popular analysis tool when understanding the performance of pattern recognition algorithms because it provides a balanced indication of both the precision and recall (Shai Shalev-Shwartz & Shai Ben-David, 2014). Precision can be understood as the fraction of times that a model is correct when it predicts that a source is present, and recall is the fraction of times that a source is present and the model "catches it". The F1 measure uses the harmonic mean of the two values, which more heavily penalizes poor performance in either metric, a consideration that becomes important when the prevalence of classes are not balanced, e.g. a class is not present much more often than it is present. The F_1 measures that were used for model evaluation were those calculated on data that was left out of the training data for both regression and classification and are shown in Table 3 for each combination of models.

$$F_{1,all} = \frac{1}{n} * \sum_{i=1:n} F1_i \tag{1}$$

$$F_{1,i} = 2 * \frac{precision_i * recall_i}{precision_i + recall_i} \tag{2}$$

$$precision_i = \frac{tp_i}{tp_i + fp_i} \tag{3}$$

$$recall_i = \frac{tp_i}{tp_i + fn_i} \tag{4}$$

[1] n indicates the number of classes (sources).

Table 3. F_1 Scores calculated for different combinations of regression and classification models. Calculated using equations (1-4) on data that was held out for validation during model training. Cells are colored by the indicated score.

	Classification Model

		Logistic_class	NeurNet_class	RandFor_class	SVMgaus_class	SVMlin_class
Regression Model	**FullLM**	0.677	0.417	0.718	0.610	0.711
	GaussProc	0.416	0.401	0.394	0.288	0.535
	NeurNet	0.596	0.496	0.690	0.569	0.596
	RandFor	0.332	0.442	0.479	0.578	0.556
	RidgeLM	0.521	0.376	0.570	0.493	0.596
	SelectLM	0.600	0.531	0.545	0.493	0.596
	StepLM	0.695	0.502	0.619	0.614	0.616

Using the criterion of maximum F_1 score, the best performance was accomplished by using the full multiple linear model (FullLM) to produce concentration estimates for a set of random forest models (RandFor_class) that estimated the sources. The performance of this pair is highlighted in Figure 9, which shows that the models were most effective at identifying gasoline vapors, and the second most successful classification was for natural gas emissions. The most common mistakes made by the model were confusion between low temperature combustion and heavy exhaust. This makes intuitive sense because the only differentiating factor in this study was the presence of NO_2 during heavy exhaust emissions, a compound that the regression models struggled to accurately predict.

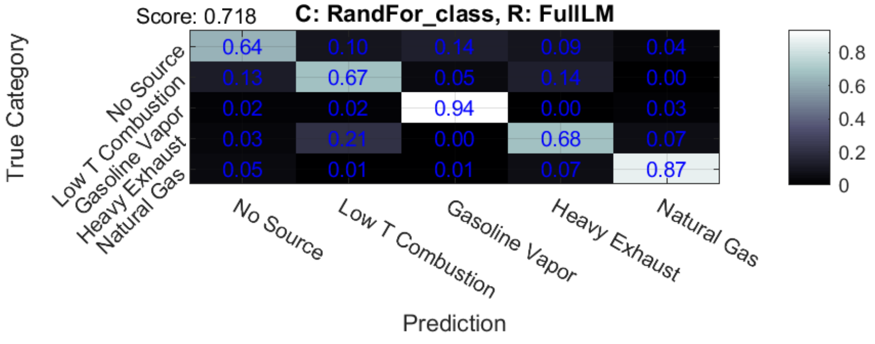


Figure 9. Confusion matrix for the "best" combination of models as judged by the F_1 score as calculated using validation data. This combination involved regression using all linear models and then random forest classification trees.

4.2. Sensor Importance for Different Compounds

The sensors that were most influential when estimating the concentration of different gases was investigated by probing parameters from a few different regression models. Specifically, we explored the sensor signals that were selected by stepwise regression, the unbiased importance measures for sensors from the random forest regression models, and the standardized ridge trace generated when fitting ridge regression models. By looking at what sensors were especially important to different models, it was possible to see which sensors were likely to be truly important to estimating the concentration of a gas.

4.2.1. Terms Selected by Stepwise Regression

Figure 10 shows the sensors selected by stepwise linear regression for each compound of interest where the boxplots indicate the statistical importance of those sensors when they were selected. Many of the terms selected algorithmically match the sensors that we have used based on previous experience (Casey et al., 2019; A. Collier-Oxandale et al., 2018), although there were a few surprises that would suggest further investigation. For example, terms that were expected were the inclusion of the Figaro 2600 and 2602 for the detection of methane and nonmethane hydrocarbons (NMHC). The importance of the Baseline Mocon sensor for gasoline was also expected, as it is one of the few sensors that our research group has used to detect heavy hydrocarbons at sub ppm levels. Finally, the sole selection of the NDIR sensor for CO_2 from NLT was expected as this technology is generally quite robust to cross-interferences and has been used successfully before in other studies. Some of the unexpected terms were the inclusion of the CO_2 sensor for indication of CO and NO_2, although this is likely because NO_2 was only ever present when CO_2 was also present, although not always at the same ratios, and the reverse was not true. CO is interesting in that there were some test points where only CO was present, although it is possible that there were not enough to counteract the more common correlation between the two compounds. It's also interesting that the gasoline models consistently selected the Figaro 4161, which is an older metal oxide sensor that was marketed for the detection of CO_2. Other sensors were selected for inclusion in only one or two of the cross-validation folds for NO_2, CO, and NMHC, which indicates that their inclusion is not reliable and may be caused by a correlation unrelated to sensitivities.

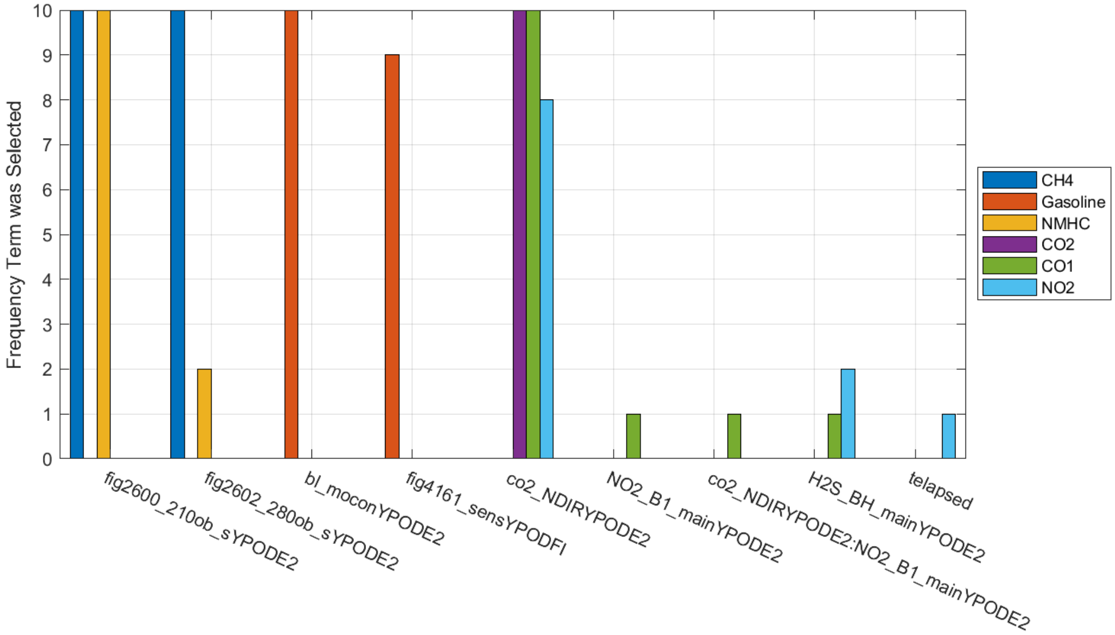


Figure 10. Frequency that each term was selected by stepwise regression for each pollutant from all cross validation folds. When an interaction term was selected, it was named with "Sensor1:Sensor2" where "Sensor1" and Sensor2" are replaced with the sensor names.

4.2.2. Random Forest Unbiased Importance Estimates

The random forests were trained such that they determined the variable to split at each node by using the interaction-curvature test as implemented by the MATLAB function "fitrtree". This method allows the model to account for interactions between variables and produces unbiased estimates of the change in mean squared error (MSE) associated with the variable selected to split at each node. Because the decision of which variable to split on is evaluated for each tree at each node using different subsets of the full set of variables, importance estimates are generated for every variable and is determined on a wide range of subsets of the data. The ten most important variables, in this case sensor signals, for each compound are shown in Figure 11.

For NO_2, we once again see that none of the sensors were selected to have a strong importance. This again is likely at least partly due to the failure of the NO_2 sensor part way through testing. This is reflected in the high ranking of the elapsed time ("telapsed") variable, as it would help separate the point of failure for that sensor. The H_2S electrochemical sensor is again among the more important inputs, reinforcing the hypothesis that it is at least partially sensitive to NO_2. Interestingly, the NO_2 sensor does feature prominently in the CO importance plot. It is possible that the individual trees making up the random forest learned to screen the NO_2 sensor when it had failed so as to be able to still use its common association with NO_2, although it is not clear why that would be true when estimating CO and not NO_2.

Both the methane and nonmethane hydrocarbon (NMHC) models are dominated by the Figaro 2600, which is the sensor that we have shown to be effective for similar applications (Casey et al., 2019; A. Collier-Oxandale et al., 2018). It is possible that the NMHC are simply taking advantage of the partial correlation of NMHC with methane and not actually recording a sensitivity of the Figaro 2600 to other NMHC. This is reinforced by the lower ranking of the 2600 when regressing gasoline concentrations. In this model, the Baseline Mocon PID sensor and Figaro 2602 are selected, and their selection matches our experience with detecting heavier hydrocarbons. The elapsed time variable is relatively important here, which matches a drift in the Figaro 2602 that was noticed when reviewing the sensor values over the full test period.

The CO_2 model is also dominated by the NDIR CO_2 sensor from ELT, another instance where the models match expectations (Casey et al., 2019; Ashley Collier-Oxandale et al., 2018). The gasoline model is also again dominated by the Baseline Mocon PID sensor and the Figaro 2602, which are two sensors that we have used to study those compounds in the past. The MICS-5121wp sensor by e2v is also relatively important, and it, too, is a sensor targeted at the detection of VOCs.

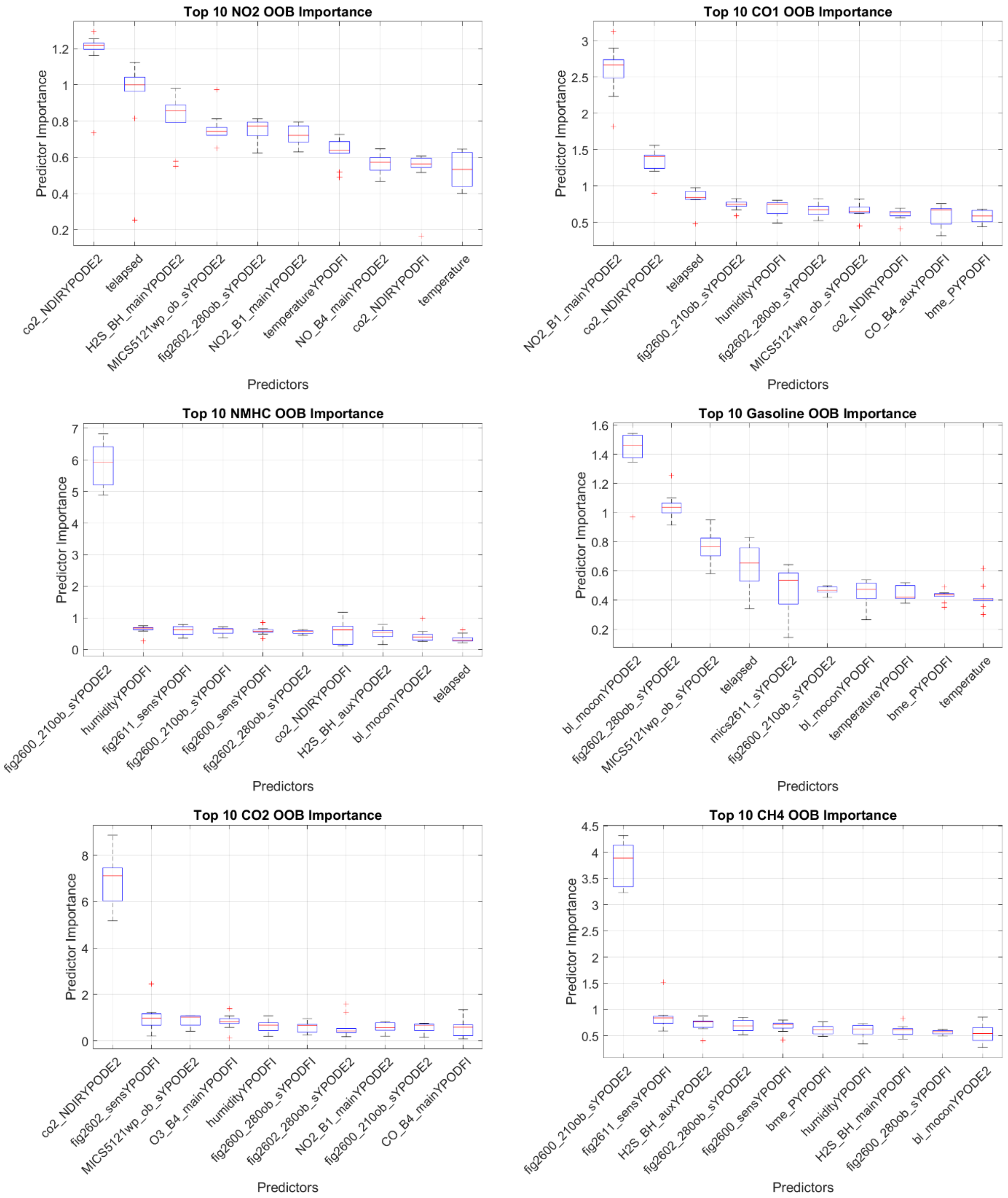


Figure 11. Box plot of the top 10 predictors sorted by out of bag importance estimates for each regressed compound. Boxes indicate the variation between cross-validation folds.

4.2.3. Standardized Ridge Regression Coefficients

The use of ridge regression offered an opportunity to review which sensors contributed to models when using regularization to avoid overfitting. Because of this regularization, more initial terms could be included; in this case interactions between each sensor and temperature, humidity, and both temperature and humidity were included in the input dataset. These are denoted by appending ":T", ":H", or ":T:H" respectively to the end of the sensor name. The methodology here initially used a range of regularization coefficients to produce a "ridge trace" plot that showed the standardized weight assigned to each variable as a function of that regularization coefficient (see Figure 12). By reviewing the ridge trace produced for each pollutant, it was possible to select a regularization coefficient that seemed to be an appropriate combination of reducing variance while avoiding underfitting. For this study, that value seemed to typically range between 1 and 10 for all pollutants, so a value of 5 was applied to all models for consistency and to avoid over-learning that hyperparameter. The ridge trace plots for each compound provide another insight into which sensors were important enough to the model that its importance to the model outweighed all but extreme values of regularization. This is indicated by weight values that are relatively high and stable, even at high levels of regularization (the right side of the plot).

In the NO_2 ridge trace, most of the weights collapsed toward zero as regularization was increased, indicating that none were especially strongly correlated with NO_2. That said, the H2S sensor was relatively resistant to regularization, indicating that it may have some sensitivity to NO_2. The CO_2 NDIR sensor term with temperature interaction term is also quite stable, although this may simply indicate that NO_2 was sometimes co-emitted with CO_2. It is important to again note that the

NO_2 sensor included in the study appears to have failed during the second half of testing, likely explaining its absence from this plot.

For CO, a strong dependence was seen with the NO_2 sensor, although this may again be caused by the fact that NO_2 was sometimes released at the same time as CO. The CO_2 NDIR sensor is also relatively resistant to regularization, but this may be caused by the same phenomena. Interestingly, the electrochemical ozone sensor (with temperature interaction) was relatively strongly and negatively represented in the ridge traces. This may indicate that the catalytic surface on the ozone sensor is somewhat sensitive to carbon monoxide, although the electrochemical cell would be operating in "reverse" as ozone is an oxidizing agent and CO is reducing.

Nonmethane hydrocarbons (NMHC) seem to have been dominated by their inclusion as a component of some of the simulated natural gas emissions, as both Figaro 2600s were strongly weighted even with high regularization. Temperature itself was also selected, although it is not clear why this was selected rather than the terms that included temperature interactions.

Gasoline vapors produced a significantly different plot from NMHC, likely because they were not co-emitted with methane and were not dominated by the ethane and propane signals from the simulated natural gas emissions. These vapors were well captured by the Baseline Mocon PID sensor that is designed to quantify gaseous hydrocarbons. Interestingly, the parameter of just this sensor and the parameter with the interaction of this sensor with temperature and humidity were both selected strongly. This may indicate that the sensor has some sensitivity to environmental parameters as well as the mixture itself. The Figaro 2602 with humidity interaction was also selected strongly, which is an expected result as these vapors are the target of the 2602 and humidity is known to affect metal oxide sensors. One of the pressure sensors (BME) was also selected with almost as much weight as the Figaro 2600 sensor. After reviewing the data, we determined that this was due to a spurious correlation with pressure in the chamber, which was not controlled.

For CO_2, several of the Figaro 2600 sensors were selected, although it is not clear whether that is representative of a true cross-sensitivity to CO_2. It is, for example, possible that these are actually cross sensitive to CO, which was co-emitted with CO2 in some of the testing. The CO2 NDIR sensor on one of the pods was selected, although not as distinctly as one might have predicted, based on the typical sensitivity of these sensors.

Finally, for methane, the Figaro 2600 sensor term was selected most strongly at higher levels of regularization. The Figaro 2600 sensor that was installed on the ancillary board and which operated at a slightly different operating temperature was also selected somewhat at a similar level to the Baseline Mocon PID sensor. Both BME pressure sensors were again selected, representing an incidental correlation with pressure.

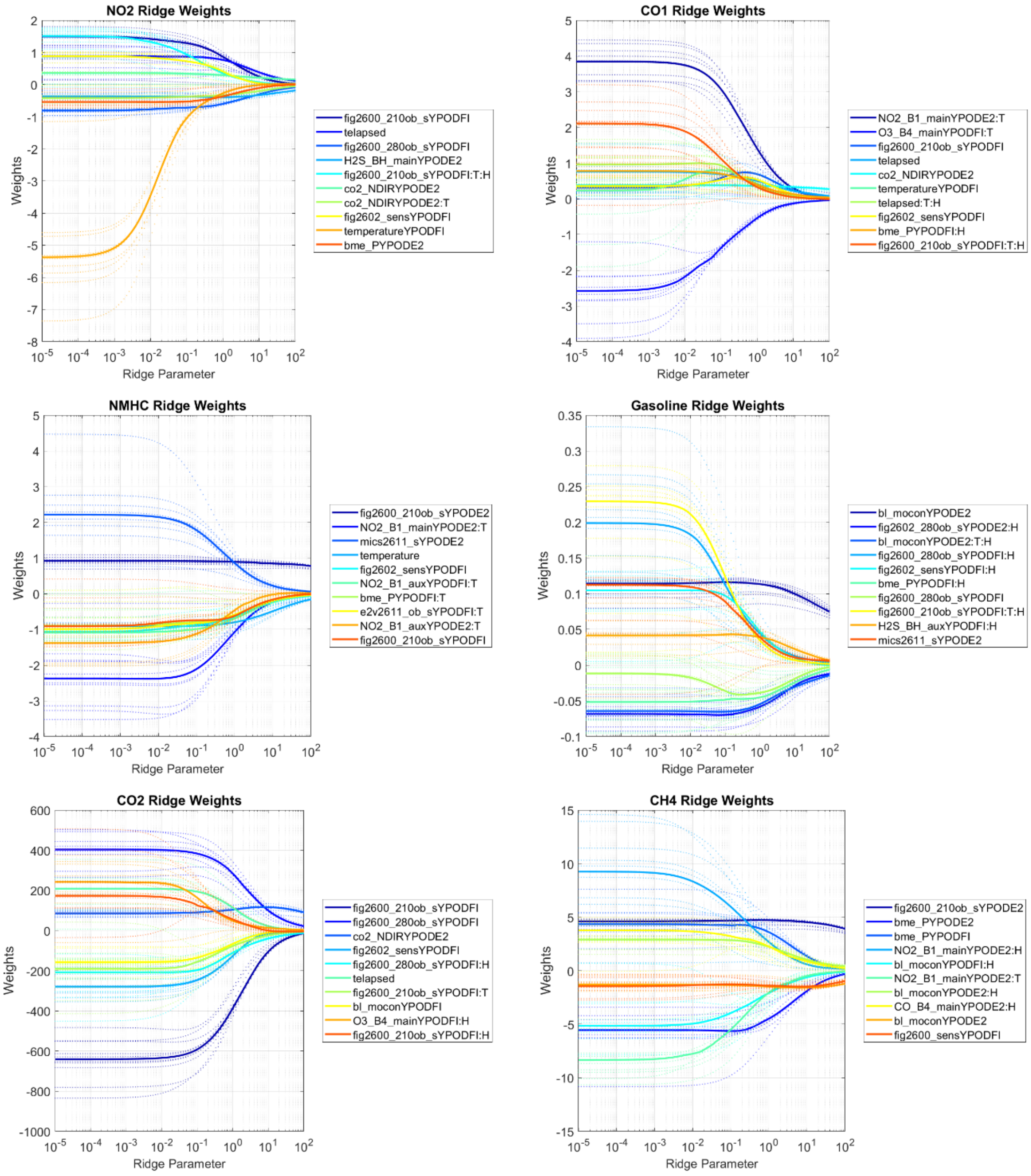


Figure 12. Summary of ridge traces of the top 10 most important parameters from the regression of each pollutant. Weight magnitude should not be compared between compounds because they scale with the concentration values. Interactions with temperature and/or humidity are indicated by ":T", ":H", or ":T:H, respectively. Colors indicate the sensor and are sorted from largest to smallest importance at high regularization. Dashed lines illustrate the ridge trace for each cross-validation fold, and the solid lines illustrate the median value for all folds. The x-axis of each plot indicates the ridge parameter (regularization strength), and the y-axis indicates the unbiased parameter weight. Larger absolute values indicate a stronger importance of that sensor, especially at larger values of the ridge parameter.

5. Conclusions

In this study, we have shown the ability of current modeling and machine learning techniques to quantify several trace gases and identify likely sources for those trace gases using an array of low-cost sensors. The best combination of regression and classification methods for source identification was a linear regression using every sensor in the array followed by a random forest of classification trees. The F_1 score as calculated on data that was held out during optimization and model fitting was 0.718, indicating a relatively strong combination of accuracy and precision. The selection of these models was a somewhat unexpected result as it was expected that the baseline linear regression model with every sensor was likely to overfit while also being unlikely to capture interactions caused by gas interferences and environmental parameters. Indeed, as far as regression accuracy, many of the other regression models outperformed this model when predicting on validation data. That said, it is possible that the random forest classifier model was able to encode the interferences that would have been normally captured by the regression model while also taking advantage of signals that would have been removed or reduced by other regression models.

In this study, we also reviewed the variable importance from three different regression models to investigate the value of sensors when quantifying different pollutants of interest. Although there were some issues caused by sensor failure, these may provide valuable insight to researchers designing the next generation of sensor platforms.

Going forward, it will be important to validate these methods in more realistic environments as chamber calibrations are not ideal representations of field performance (Masson, Piedrahita, & Hannigan, 2015; Piedrahita et al., 2014b). One example of a study that could provide valuable insights would be the placement of this sensor array in regions where different sources are likely to dominate, e.g. immediately next to a highway onramp and then next to a well during completion. By collecting data in these locations as well as e.g. rural sites, it would be possible to determine how well the identification of sources might function in situations where the qualitative levels are known. This has been demonstrated by Molino and others for the detection of single sources (Molino et al., 2012).

There are also other analysis techniques that could be used to possibly improve the results here. Specifically, combining the models into ensembles for both classification and regression, the ensembles could improve overall results by combining models that make "decisions" in different ways and are therefore less likely to make the same types of mistakes. The use of generative classification models could also improve the identification of sources. These models learn the characteristics of different classes instead of learning to separate one class from another. This type of classifier generally performs better on smaller training data sets, and relative to the possible conditions that sensors could be exposed to in field campaigns, any calibration will always represent a small training set.

Finally, many sensors were not available for inclusion in this array due to limited availability and schedule conflicts. Including new sensors in the array could fill in gaps that were not covered by the relatively limited variety of sensors here. These new inputs could involve similar sensors operated in new ways, e.g. temperature cycled MOx sensors that have been shown to improve sensitivity to VOCs and other gases (Lee & Reedy, 1999; Schüler et al., 2014; Schütze et al., 2017). It could also include sensors with wildly different operating principles like micro-GCs (Spinelle, Gerboles, Kok, et al., 2017) or with new targets like the low-cost PM sensors that could also be useful in the differentiation of sources.

6. REFERENCES

Airgas. (2018, February). Specialty Gases and Equipment Product Reference Guide. Retrieved from http://airgassgcatalog.com/catalog/

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., … Wennberg, P. O. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, *11*(9), 4039–4072. https://doi.org/10.5194/acp-11-4039-2011

Akamatsu, T. I. T., Tsuruta, A., & Shin, W. (2017). Selective Detection of Target Volatile Organic Compounds in Contaminated Humid Air Using a Sensor Array with Principal Component Analysis. *Sensors*, *17*(7), 1662. https://doi.org/10.3390/s17071662

Aleixandre, M., Sayago, I., Horrillo, M. C., Fernández, M. J., Arés, L., García, M., … Gutiérrez, J. (2004). Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas. *Sensors and Actuators B: Chemical*, *103*(1), 122–128. https://doi.org/10.1016/j.snb.2004.04.044

Alphasense Ltd. (2009, March). AAN 105-03: Designing a Potentiostatic Circuit. Alphasense Ltd. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_105-03.pdf

Alphasense Ltd. (2013). AAN 109-02: Interfering Gases. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_109-02.pdf

Bastuck, M., Bur, C., Lloyd Spetz, A., Andersson, M., & Schütze, A. (2014). Gas identification based on bias induced hysteresis of a gas-sensitive SiC field effect transistor. *Journal of Sensors and Sensor Systems*, *3*(1), 9–19. https://doi.org/10.5194/jsss-3-9-2014

Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G., & Hueglin, C. (2018). Performance of NO, NO2 low cost sensors and three calibration approaches within a real world application. *Atmospheric Measurement Techniques*, *11*(6), 3717–3735. https://doi.org/10.5194/amt-11-3717-2018

Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., … Penza, M. (2018). Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II. *Atmospheric Environment*, *193*, 127–142. https://doi.org/10.1016/j.atmosenv.2018.08.028

Capone, S., Siciliano, P., Bârsan, N., Weimar, U., & Vasanelli, L. (2001). Analysis of CO and CH4 gas mixtures by using a micromachined sensor array. *Sensors and Actuators B: Chemical*, *78*(1), 40–48. https://doi.org/10.1016/S0925-4005(01)00789-4

Casey, J. G., Collier-Oxandale, A., & Hannigan, M. P. (2019). Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. *Sensors and Actuators B: Chemical*, *283*, 504–514. https://doi.org/10.1016/j.snb.2018.12.049

Cheadle, L., Deanes, L., Sadighi, K., Casey, J. G., Collier-Oxandale, A., & Hannigan, M. (2017). Quantifying Neighborhood-Scale Spatial Variations of Ozone at Open Space and Urban Sites in Boulder, Colorado Using Low-Cost Sensor Technology. *Sensors; Basel*, *17*(9), 2072. http://dx.doi.org.colorado.idm.oclc.org/10.3390/s17092072

Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., … Hannigan, M. (2017). Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). *Sensors*, *17*(11), 2478. https://doi.org/10.3390/s17112478

Collier-Oxandale, A., Hannigan, M. P., Casey, J. G., Piedrahita, R., Ortega, J., Halliday, H., & Johnston, J. (2018). Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. *Atmos. Meas. Tech. Discuss.*, *2018*, 1–35. https://doi.org/10.5194/amt-2017-421

Collier-Oxandale, Ashley, Coffey, E., Thorson, J., Johnston, J., & Hannigan, M. (2018). Comparing Building and Neighborhood-Scale Variability of CO2 and O3 to Inform Deployment Considerations for Low-Cost Sensor System Use. *Sensors (Basel, Switzerland)*, *18*(5). https://doi.org/10.3390/s18051349

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., … Jayne, J. T. (2017). Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. *Atmos. Meas. Tech.*, *10*(9), 3575–3588. https://doi.org/10.5194/amt-10-3575-2017

De Vito, S., Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., & Di Francia, G. (2018). Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches. *Sensors and Actuators B: Chemical*, *255*, 1191–1210. https://doi.org/10.1016/j.snb.2017.07.155

De Vito, Saverio, Castaldo, A., Loffredo, F., Massera, E., Polichetti, T., Nasti, I., … Di Francia, G. (2007). Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures. *Sensors and Actuators B: Chemical*, *124*(2), 309–316. https://doi.org/10.1016/j.snb.2006.12.039

De Vito, Saverio, Piga, M., Martinotto, L., & Di Francia, G. (2009). CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators B: Chemical*, *143*(1), 182–191. https://doi.org/10.1016/j.snb.2009.08.041

Fonollosa, J., Sheik, S., Huerta, R., & Marco, S. (2015). Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. *Sensors and Actuators B: Chemical*, *215*, 618–629. https://doi.org/10.1016/j.snb.2015.03.028

Garcia, C. AB-617 Nonvehicular air pollution: criteria air pollutants and toxic air contaminants., Chapter 136 California Health and Safety Code § 40920.6, 42400, 42402, 39607.1, 40920.8, 42411, 42705.5, and 44391.2 (2017). Retrieved from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB617

Hagler, G. S. W., Williams, R., Papapostolou, V., & Polidori, A. (2018). Air Quality Sensors and Data Adjustment Algorithms: When Is It No Longer a Measurement? *Environmental Science & Technology*, *52*(10), 5530–5531. https://doi.org/10.1021/acs.est.8b01826

Heimann, I., Bright, V. B., McLeod, M. W., Mead, M. I., Popoola, O. A. M., Stewart, G. B., & Jones, R. L. (2015). Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. *Atmospheric Environment*, *113*(Supplement C), 10–19. https://doi.org/10.1016/j.atmosenv.2015.04.057

Kim, J., Shusterman, A. A., Lieschke, K. J., Newman, C., & Cohen, R. C. (2017). The BErkeley Atmospheric CO2 Observation Network: Field Calibration and Evaluation of Low-cost Air Quality Sensors. *Atmos. Meas. Tech. Discuss.*, *2017*, 1–20. https://doi.org/10.5194/amt-2017-339

Korotcenkov, G., & Cho, B. K. (2011). Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey). *Sensors and Actuators B: Chemical*, *156*(2), 527–538. https://doi.org/10.1016/j.snb.2011.02.024

Korotcenkov, Ghenadii. (2013). *Handbook of Gas Sensor Materials*. (Ghenadii Korotcenkov, Ed.) (Vol. 1: Conventional Approaches). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7165-3_11

Lee, A. P., & Reedy, B. J. (1999). Temperature modulation in semiconductor gas sensing. *Sensors and Actuators B: Chemical*, *60*(1), 35–42. https://doi.org/10.1016/S0925-4005(99)00241-5

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., … White, A. (2016a). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions*, *189*(0), 85–103. https://doi.org/10.1039/C5FD00201J

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., … White, A. (2016b). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions*, *189*(0), 85–103. https://doi.org/10.1039/C5FD00201J

Marco, S., & Gutierrez-Galvez, A. (2012). Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review. *IEEE Sensors Journal*, *12*(11), 3189–3214. https://doi.org/10.1109/JSEN.2012.2192920

Martenies, S. E., Milando, C. W., Williams, G. O., & Batterman, S. A. (2017). Disease and Health Inequalities Attributable to Air Pollutant Exposure in Detroit, Michigan. *International Journal of Environmental Research and Public Health*, *14*(10), 1243. https://doi.org/10.3390/ijerph14101243

Masson, N., Piedrahita, R., & Hannigan, M. (2015). Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring. *Sensors and Actuators B: Chemical*, *208*, 339–345. https://doi.org/10.1016/j.snb.2014.11.032

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., … Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, *70*(Supplement C), 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060

Miskell, G., Salmond, J., Grange, S., Weissert, L., Henshaw, G., & Williams, D. (2017). Reliable Long-Term Data from Low-Cost Gas Sensor Networks in the Environment. *Proceedings*, *1*(4), 400. https://doi.org/10.3390/proceedings1040400

Molino, A., Elen, B., Theunis, J., Ingarra, S., Van den Bossche, J., Reggente, M., & Loreto, V. (2012). The EveryAware SensorBox: a tool for community-based air quality monitoring. In *Workshop Sensing a Changing World*.

Morel, P. (2018). Gramm: grammar of graphics plotting in Matlab. *The Journal of Open Source Software*. https://doi.org/10.21105/joss.00568

Penza, M., Suriano, D., Pfister, V., Prato, M., & Cassano, G. (2017). Urban Air Quality Monitoring with Networked Low-Cost Sensor-Systems. *Proceedings*, *1*(4), 573. https://doi.org/10.3390/proceedings1040573

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., … Shang, L. (2014a). The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques; Katlenburg-Lindau*, *7*(10), 3325. http://dx.doi.org/10.5194/amt-7-3325-2014

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., … Shang, L. (2014b). The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques; Katlenburg-Lindau*, *7*(10), 3325. http://dx.doi.org/10.5194/amt-7-3325-2014

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., … Jones, R. L. (2018). Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmospheric Environment*, *194*, 58–70. https://doi.org/10.1016/j.atmosenv.2018.09.030

Robinson, J. A., Kocman, D., Horvat, M., & Bartonova, A. (2018). End-User Feedback on a Low-Cost Portable Air Quality Sensor System—Are We There Yet? *Sensors*, *18*(11), 3768. https://doi.org/10.3390/s18113768

Romain, A. C., & Nicolas, J. (2010). Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. *Sensors and Actuators B: Chemical*, *146*(2), 502–506. https://doi.org/10.1016/j.snb.2009.12.027

Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D. K., & Hannigan, M. (2018). Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors. *Atmospheric Measurement Techniques*, *11*(3), 1777–1792. https://doi.org/10.5194/amt-11-1777-2018

Schüler, M., Fricke, T., Sauerwald, T., & Schütze, A. (2015). E8.4 - Detecting poisoning of metal oxide gas sensors at an early stage by temperature cycled operation. *Proceedings SENSOR 2015*, 735–739. http://dx.doi.org/10.5162/sensor2015/E8.4

Schüler, M., Sauerwald, T., & Schütze, A. (2014). Metal oxide semiconductor gas sensor self-test using Fourier-based impedance spectroscopy. *Journal of Sensors and Sensor Systems*, *3*(2), 213–221. https://doi.org/10.5194/jsss-3-213-2014

Schultealbert, C., Baur, T., Schütze, A., & Sauerwald, T. (2018). Facile Quantification and Identification Techniques for Reducing Gases over a Wide Concentration Range Using a MOS Sensor in Temperature-Cycled Operation. *Sensors*, *18*(3), 744. https://doi.org/10.3390/s18030744

Schütze, A., Baur, T., Leidinger, M., Reimringer, W., Jung, R., Conrad, T., & Sauerwald, T. (2017). Highly Sensitive and Selective VOC Sensor Systems Based on Semiconductor Gas Sensors: How to? *Environments*, *4*(1), 20. https://doi.org/10.3390/environments4010020

Shai Shalev-Shwartz, & Shai Ben-David. (2014). *Understanding Machine Learning: From Theory to Algorithms*. New York, NY: Cambridge University Press. Retrieved from http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/

Shamasunder, B., Collier-Oxandale, A., Blickley, J., Sadd, J., Chan, M., Navarro, S., … Wong, N. J. (2018). Community-Based Health and Exposure Study around Urban Oil Developments in South Los Angeles. *International Journal of Environmental Research and Public Health*, *15*(1), 138. https://doi.org/10.3390/ijerph15010138

Spinelle, L., Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017). Performance Evaluation of Low-Cost BTEX Sensors and Devices within the EURAMET Key-VOCs Project. *Proceedings*, *1*(4), 425. https://doi.org/10.3390/proceedings1040425

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2. *Sensors and Actuators B: Chemical*, *238*(Supplement C), 706–715. https://doi.org/10.1016/j.snb.2016.07.036

Suárez, J. I., Arroyo, P., Lozano, J., Herrero, J. L., & Padilla, M. (2018). Bluetooth gas sensing module combined with smartphones for air quality monitoring. *Chemosphere*, *205*, 618–626. https://doi.org/10.1016/j.chemosphere.2018.04.154

Szczurek, A., Szecówka, P. M., & Licznerski, B. W. (1999). Application of sensor array and neural networks for quantification of organic solvent vapours in air. *Sensors and Actuators B: Chemical*, *58*(1), 427–432. https://doi.org/10.1016/S0925-4005(99)00105-7

Tan, Y., Lipsky, E. M., Saleh, R., Robinson, A. L., & Presto, A. A. (2014). Characterizing the Spatial Variation of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA. *Environmental Science & Technology*, *48*(24), 14186–14194. https://doi.org/10.1021/es5034074

Tangirala, V. K. K., Gómez-Pozos, H., Rodríguez-Lugo, V., & Olvera, M. D. L. L. (2017). A Study of the CO Sensing Responses of Cu-, Pt- and Pd-Activated SnO2 Sensors: Effect of Precipitation Agents, Dopants and Doping Methods. *Sensors (Basel, Switzerland)*, *17*(5). https://doi.org/10.3390/s17051011

Tomchenko, A. A., Harmer, G. P., Marquis, B. T., & Allen, J. W. (2003). Semiconducting metal oxide sensor array for the selective detection of combustion gases. *Sensors and Actuators B: Chemical*, *93*(1), 126–134. https://doi.org/10.1016/S0925-4005(03)00240-5

Tsujita, W., Yoshino, A., Ishida, H., & Moriizumi, T. (2005). Gas sensor network for air-pollution monitoring. *Sensors and Actuators B: Chemical*, *110*(2), 304–311. https://doi.org/10.1016/j.snb.2005.02.008

US EPA, O. (2014, August 19). Integrated Urban Air Toxics Strategy [Reports and Assessments]. Retrieved September 25, 2018, from https://www.epa.gov/urban-air-toxics/integrated-urban-air-toxics-strategy

US EPA, O. (2016). *2014 National Emission Inventory (NEI) Report* (Policies and Guidance). Retrieved from https://www.epa.gov/air-emissions-inventories/2014-national-emission-inventory-nei-report

Vembu, S., Vergara, A., Muezzinoglu, M. K., & Huerta, R. (2012). On time series features and kernels for machine olfaction. *Sensors and Actuators B: Chemical*, *174*, 535–546. https://doi.org/10.1016/j.snb.2012.06.070

Vergara, A., Llobet, E., Brezmes, J., Ivanov, P., Vilanova, X., Gracia, I., … Correig, X. (2005). Optimised temperature modulation of metal oxide micro-hotplate gas sensors through multilevel pseudo random sequences. *Sensors and Actuators B: Chemical*, *111*(Supplement C), 271–280. https://doi.org/10.1016/j.snb.2005.06.039

Westerholm, R., & Egebäck, K. E. (1994). Exhaust emissions from light- and heavy-duty vehicles: chemical composition, impact of exhaust after treatment, and fuel parameters. *Environmental Health Perspectives*, *102*(Suppl 4), 13–23.

Woodall, G. M., Hoover, M. D., Williams, R., Benedict, K., Harper, M., Soo, J.-C., … Thompson, B. (2017). Interpreting Mobile and Handheld Air Sensor Readings in Relation to Air Quality Standards and Health Effect Reference Values: Tackling the Challenges. *Atmosphere*, *8*(10), 182. https://doi.org/10.3390/atmos8100182

Yokelson, R. J., Burling, I. R., Gilman, J. B., Warneke, C., Stockwell, C. E., Gouw, J. de, … Weise, D. R. (2013). Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires. *Atmospheric Chemistry and Physics*, *13*(1), 89–116. https://doi.org/10.5194/acp-13-89-2013

Yoo, K. S. (2011). Gas Sensors for Monitoring Air Pollution. https://doi.org/10.5772/19967

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., … Subramanian, R. (2017). Closing the gap on lower cost air quality monitoring: machine learning calibration models to improve low-cost sensor performance. *Atmos. Meas. Tech. Discuss.*, *2017*, 1–36. https://doi.org/10.5194/amt-2017-260

Zimmerman, Naomi, Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., … Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques; Katlenburg-Lindau*, *11*(1), 291–313. http://dx.doi.org/10.5194/amt-11-291-2018

CHAPTER 3

CONCLUSIONS

This thesis has focused on the many sources of VOCs and the development of a tool to identify those sources at a more localized scale than is currently feasible. The first chapter presented an overview of some of the major sources of VOCs that affect air quality across the United States. That chapter provided a relatively brief overview of those sources, including some of the compounds that are currently understood to be commonly associated with them. The variation of chemical composition both within and between VOC sources quickly becomes apparent, motivating the studies that have been undertaken to better understand their contribution to ambient air quality. I then reviewed some of the methodologies underlying those studies in order to understand their effectiveness. Using this background, I identified the gaps that the need to be addressed in order to understand the sources responsible for VOC exposures at a community scale. The development of a low-cost sensor array that is able to fill some of these gaps was the focus of the research presented in Chapter two. Before discussing their application in Chapter 2, the underlying sensor technologies were briefly reviewed in order to give the reader a better understanding of the diversity of technologies that were applied in my research.

One benefit of the use of low-cost sensor tools is making air quality information more accessible to a broad range of users. By reducing the costs of sensors by orders of magnitude, developers have helped to reduce the financial barriers to the adoption of these research tools for individuals and communities that do not have the financial resources to support a study using traditional research tools. However, collecting data is only the first step in these studies. Typically, users deploy low-cost sensors with a specific goal in mind, and that goal requires an understanding of the outputs of those sensors. In conversations with low-cost sensor users, they have described many of their own goals. Those goals included, for example, deciding when to collect limited grab samples, collecting data that may persuade local government to take an action, or understanding whether a specific oil and gas development is affecting their air quality.

Identifying the sources of measured compounds is one of the more challenging tasks when studying air quality, whether it be with research grade instruments or low-cost sensors. It becomes even more difficult when that research is conducted with low-cost sensors because of poorly understood cross-sensitivities to other gases, temperature and humidity effects, temporal drift, and high noise. In the course of the research presented in Chapter 2, I was able to use several tools to better understand some of these effects and advance the state of knowledge. As we discussed in Chapter 1, there are a variety of processes that can produce VOCs, so measurements of the concentration of one or two compounds is rarely enough to answer those questions. However, measurement campaigns that only capture a few compounds and attempt to determine sources from those measurements represent the bulk of the current literature for low-cost sensors. In Chapter 2, I developed an array of low-cost sensors and complementary analysis tools to address the task of both quantifying the concentration of gaseous pollutants and algorithmically estimating the presence of sources that could have emitted the detected gases.

The research presented in the second chapter was intended as an initial investigation of the feasibility of algorithmically identifying multiple sources, something that has not been attempted in the literature, as far as I am aware. In impact the entire United States as well as the communities that I have worked with in Southern California. Understanding the major components of those sources allowed me to identify what sensors and regression methods had been successfully applied in the literature to detecting those compounds.

After this preliminary review, I developed a test matrix that included the simulation of four sources: biomass burning, mobile sources, unburned liquid fossil fuel emissions, and unburned natural gas emissions. These sources were selected based on their presence in the literature and their applicability to community groups that we have worked with in Southern California. By choosing to conduct this first study in our test chamber, I was able to reduce correlations between different compounds, environmental parameters, and time. I accomplished this by varying the exact mixtures that made up each simulated source and the temperature and humidity conditions for each of those mixtures. This design presented a dataset designed to make it less likely for models to learn, for example, that carbon monoxide was always emitted with NO2, or that concentrations were correlated with diurnal temperature changes. This contributed to the state of knowledge by creating a dataset with which it was easier to understand how each sensor in the array reacts to variation in individual parameters like concentration or temperature. Test points were also repeated to allow for some understanding of which sensors were sensitive to a temporal drift.

While conducting the chamber experiments, I was also creating a MATLAB script that can repeatably import and filter the sensor data, apply regression models, and then apply classification models to use those regressed concentration estimates to identify sources. This two-step methodology that I used to directly estimate the likely sources of measured pollutants is novel, as far as I am aware. Because of the novelty of this approach, I applied multiple classification techniques that were selected because they operate in significantly different ways. By publishing the performance of several classification models this research will inform future studies with similar goals or that use classification to understand other aspects of air quality.

The accuracies that we achieved while regressing specific compounds were acceptable, although researchers including those in the Hannigan Lab have achieved accuracies higher than were achieved here during campaigns targeted at quantifying individual compounds. The significant result here was that the classification models were still able to use those estimates to consistently identify the source that was being simulated.

Although this success is encouraging, the scope of the chamber study was intentionally limited and was conducted entirely in a laboratory setting. For a researcher looking to further this work, I present here a few suggestions of improvements that seem appropriate with the benefit of the knowledge gathered here. Primarily, I would suggest collaborating with sensor manufacturers if at all possible. Greater involvement of sensor manufacturers would be beneficial to this study because a more fundamental understanding of the differences and similarities of sensors would inform the selection of which sensors to include in the array. Any agreement would, need to protect the intellectual property of sensor manufacturers, but I believe that it would be possible to incorporate this knowledge without revealing their intellectual property. Greater communication of the analytical models and methods used by researchers in the low-cost sensor space would be greatly beneficial to the state of knowledge. Although most researchers in this space are not computer science or machine learning experts, the communication of the specific parameters that were used in models would greatly improve reproducibility and would contribute to our understanding of the efficacy of different modeling techniques as applied to low-cost sensors – a topic that is still highly debated and not well understood.

In addition to those more general improvements, the concrete next step in this research will be to expose the sensor array to more realistic deployments. That may include additional chamber studies that simulate multiple pollutants simultaneously or field deployments in areas that have a single dominant source. It will also be important to study the use of a system like this in a realistic interaction with potential users. Measurements of pollutant concentrations from low-cost sensors are of little use unless they are understood by the users, and the usefulness of an automated source identification be limited if it is not presented in a way that is clearly understood by the end users.

BIBLIOGRPHY

Aires, F., Prigent, C., & Rossow, W. B. (2004). Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 2. Output errors. Journal of Geophysical Research: Atmospheres, 109(D10), D10304. https://doi.org/10.1029/2003JD004174

Airgas. (2018, February). Specialty Gases and Equipment Product Reference Guide. Retrieved from http://airgassgcatalog.com/catalog/

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., … Wennberg, P. O. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. Atmospheric Chemistry and Physics, 11(9), 4039–4072. https://doi.org/10.5194/acp-11-4039-2011

Akamatsu, T. I. T., Tsuruta, A., & Shin, W. (2017). Selective Detection of Target Volatile Organic Compounds in Contaminated Humid Air Using a Sensor Array with Principal Component Analysis. Sensors, 17(7), 1662. https://doi.org/10.3390/s17071662

Alavi-Shoshtari, M., Salmond, J. A., Giurcăneanu, C. D., Miskell, G., Weissert, L., & Williams, D. E. (2018). Automated data scanning for dense networks of low-cost air quality instruments: Detection and differentiation of instrumental error and local to regional scale environmental abnormalities. Environmental Modelling & Software, 101, 34–50. https://doi.org/10.1016/j.envsoft.2017.12.002

Aleixandre, M., Sayago, I., Horrillo, M. C., Fernández, M. J., Arés, L., García, M., … Gutiérrez, J. (2004). Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas. Sensors and Actuators B: Chemical, 103(1), 122–128. https://doi.org/10.1016/j.snb.2004.04.044

Aleixandre, Manuel, & Gerboles, M. (2012). Review of Small Commercial Sensors for Indicative Monitoring of Ambient Gas. Chemical Engineering Transactions, 30. (JRC75718).

Alhasa, K. M., Mohd Nadzir, M. S., Olalekan, P., Latif, M. T., Yusup, Y., Iqbal Faruque, M. R., … Ezani, N. E. (2018). Calibration Model of a Low-Cost Air Quality Sensor Using an Adaptive Neuro-Fuzzy Inference System. Sensors, 18(12), 4380. https://doi.org/10.3390/s18124380

Allouch, A., Le Calvé, S., & Serra, C. A. (2013). Portable, miniature, fast and high sensitive real-time analyzers: BTEX detection. Sensors and Actuators B: Chemical, 182, 446–452. https://doi.org/10.1016/j.snb.2013.03.010

Almand-Hunter, B. B., Collier, A. M., Piedrahita, R. A., & Hannigan, M. P. (2015, August 6). Improving Low-Cost Metal-Oxide Sensor Calibrations with Machine Learning.

Alpert, P., Archibald, A., Arnold, S., Ashworth, K., Brown, S., Campbell, S., … Zelenyuk, A. (2017). New tools for atmospheric chemistry: general discussion. Faraday Discussions, 200(0), 663–691. https://doi.org/10.1039/C7FD90041D

Alphasense Ltd. (2009, March). AAN 105-03: Designing a Potentiostatic Circuit. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_105-03.pdf

Alphasense Ltd. (2013a). AAN 104: How Electrochemical Gas Sensors Work. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_104.pdf

Alphasense Ltd. (2013b). AAN 109-02: Interfering Gases. Retrieved from http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_109-02.pdf

Alvarez, Ramón A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., … Hamburg, S. P. (2018a). Assessment of methane emissions from the U.S. oil and gas supply chain. Science, eaar7204. https://doi.org/10.1126/science.aar7204

Alvarez, Ramón A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., … Hamburg, S. P. (2018b). Assessment of methane emissions from the U.S. oil and gas supply chain. Science, 361(6398), 186–188. https://doi.org/10.1126/science.aar7204

Alvarez, Ramon A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., … Hamburg, S. P. (2018). Improved characterization of methane emissions from the U.S. oil and gas supply chain. Science Magazine. http://dx.doi.org/10.1002/https://dx.doi.org/10.1126/science.aar7204

Aoki, P. M., Honicky, R. J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S., & Woodruff, A. (2010). Common Sense: Mobile Environmental Sensing Platforms to Support Community Action and Citizen Science.

Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. Atmospheric Environment, 34(12), 2063–2101. https://doi.org/10.1016/S1352-2310(99)00460-4

Atkinson, R., & Arey, J. (2003). Atmospheric Degradation of Volatile Organic Compounds. Chemical Reviews, 103(12), 4605–4638. https://doi.org/10.1021/cr0206420

Baker, A. K., Beyersdorf, A. J., Doezema, L. A., Katzenstein, A., Meinardi, S., Simpson, I. J., … Sherwood Rowland, F. (2008). Measurements of nonmethane hydrocarbons in 28 United States cities. Atmospheric Environment, 42(1), 170–182. https://doi.org/10.1016/j.atmosenv.2007.09.007

Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., … Pierce, T. (2016). Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches. Atmospheric Environment, 140, 539–554. https://doi.org/10.1016/j.atmosenv.2016.06.032

Barakeh, Z. A., Breuil, P., Redon, N., Pijolat, C., Locoge, N., & Viricelle, J.-P. (2017). Development of a normalized multi-sensors system for low cost on-line atmospheric pollution detection. Sensors and Actuators B: Chemical, 241, 1235–1243. https://doi.org/10.1016/j.snb.2016.10.006

Barcelo-Ordinas, J. M., Garcia-Vidal, J., Doudou, M., Rodrigo-Muñoz, S., & Cerezo-Llavero, A. (2018). Calibrating low-cost air quality sensors using multiple arrays of sensors. 2018 IEEE Wireless Communications and Networking Conference (WCNC), 1–6. https://doi.org/10.1109/WCNC.2018.8377051

Baron, R., & Saffell, J. (2017). Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review. ACS Sensors, 2(11), 1553–1566. https://doi.org/10.1021/acssensors.7b00620

Barsan, N., & Weimar, U. (2001). Conduction Model of Metal Oxide Gas Sensors. Journal of Electroceramics, 7(3), 143–167. https://doi.org/10.1023/A:1014405811371

Bastuck, M., Baur, T., Richter, M., Mull, B., Schütze, A., & Sauerwald, T. (2018). Comparison of ppb-level gas measurements with a metal-oxide semiconductor gas sensor in two independent laboratories. Sensors and Actuators B: Chemical. https://doi.org/10.1016/j.snb.2018.06.097

Bastuck, M., Bur, C., Lloyd Spetz, A., Andersson, M., & Schütze, A. (2014). Gas identification based on bias induced hysteresis of a gas-sensitive SiC field effect transistor. Journal of Sensors and Sensor Systems, 3(1), 9–19. https://doi.org/10.5194/jsss-3-9-2014

Bastuck, M., Bur, C., Sauerwald, T., Schütze, A., Spetz, A. L., & Andersson, M. (2015). D5.1 - Quantification of Volatile Organic Compounds in the ppb-range using

Partial Least Squares Regression. Proceedings SENSOR 2015, 584–589. http://dx.doi.org/10.5162/sensor2015/D5.1

Bastuck, M., Reimringer, W., Conrad, T., & Schütze, A. (2016a). Dynamic Multisensor Operation and Read-out for Highly Selective Gas Sensor Systems. Procedia Engineering, 168(Supplement C), 1685–1688. https://doi.org/10.1016/j.proeng.2016.11.490

Bastuck, M., Reimringer, W., Conrad, T., & Schütze, A. (2016b). Dynamic Multisensor Operation and Read-out for Highly Selective Gas Sensor Systems. Procedia Engineering, 168, 1685–1688. https://doi.org/10.1016/j.proeng.2016.11.490

Bastuck, Manuel, Leidinger, M., Sauerwald, T., & Schütze, A. (2015). Improved quantification of naphthalene using non-linear Partial Least Squares Regression. ArXiv:1507.05834 [Stat]. Retrieved from http://arxiv.org/abs/1507.05834

Batterman, S., Chambliss, S., & Isakov, V. (2014). Spatial resolution requirements for traffic-related air pollutant exposure evaluations. Atmospheric Environment, 94, 518–528. https://doi.org/10.1016/j.atmosenv.2014.05.065

Baur, T., Schultealbert, C., Schütze, A., & Sauerwald, T. (2018). Novel method for the detection of short trace gas pulses with metal oxide semiconductor gas sensors. Journal of Sensors and Sensor Systems, 7(1), 411–419. https://doi.org/10.5194/jsss-7-411-2018

Benammar, M., & Maskell, W. C. (1989). Temperature control of thick-film printed heaters. Journal of Physics E: Scientific Instruments, 22(11), 933. https://doi.org/10.1088/0022-3735/22/11/006

Bennetts, V. H., Schaffernicht, E., Pomareda, V., Lilienthal, A. J., Marco, S., & Trincavelli, M. (2014). Combining Non Selective Gas Sensors on a Mobile Robot for Identification and Mapping of Multiple Chemical Compounds. Sensors; Basel, 14(9), 17331–17352. http://dx.doi.org.colorado.idm.oclc.org/10.3390/s140917331

Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G., & Hueglin, C. (2018). Performance of NO, NO2 low cost sensors and three calibration approaches within a real world application. Atmospheric Measurement Techniques, 11(6), 3717–3735. https://doi.org/10.5194/amt-11-3717-2018

Borrego, C., Costa, A. M., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., … Penza, M. (2016). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. Atmospheric Environment, 147, 246–263. https://doi.org/10.1016/j.atmosenv.2016.09.050

Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., … Penza, M. (2018). Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II. Atmospheric Environment, 193, 127–142. https://doi.org/10.1016/j.atmosenv.2018.08.028

Brantley, H. L., Thoma, E. D., Squier, W. C., Guven, B. B., & Lyon, D. (2014). Assessment of Methane Emissions from Oil and Gas Production Pads using Mobile Measurements. Environmental Science & Technology, 48(24), 14508–14515. https://doi.org/10.1021/es503070q

Brattain, W. H., & Bardeen, J. (1953). Surface Properties of Germanium. Bell System Technical Journal, 32(1), 1–41. https://doi.org/10.1002/j.1538-7305.1953.tb01420.x

Burgués, J., Jiménez-Soto, J. M., & Marco, S. (2018). Estimation of the limit of detection in semiconductor gas sensors through linearized calibration models. Analytica Chimica Acta, 1013, 13–25. https://doi.org/10.1016/j.aca.2018.01.062

Burgués, J., & Marco, S. (2018). Multivariate estimation of the limit of detection by orthogonal partial least squares in temperature-modulated MOX sensors. Analytica Chimica Acta, 1019, 49–64. https://doi.org/10.1016/j.aca.2018.03.005

Burgués, J., Marco, S., Burgués, J., & Marco, S. (2018). Low Power Operation of Temperature-Modulated Metal Oxide Semiconductor Gas Sensors. Sensors, 18(2), 339. https://doi.org/10.3390/s18020339

California Air Resources Board. (2018). Community Air Protection Blueprint. Retrieved October 11, 2018, from https://ww2.arb.ca.gov/our-work/programs/community-air-protection-program/community-air-protection-blueprint

Campbell, M. G., Liu, S. F., Swager, T. M., & Dincă, M. (2015). Chemiresistive Sensor Arrays from Conductive 2D Metal–Organic Frameworks. Journal of the American Chemical Society, 137(43), 13780–13783. https://doi.org/10.1021/jacs.5b09600

Cao, L. J., & Chong, W. K. (2002). Feature extraction in support vector machine: a comparison of PCA, XPCA and ICA. Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02, 2, 1001–1005 vol.2. https://doi.org/10.1109/ICONIP.2002.1198211

Capone, S., Siciliano, P., Bârsan, N., Weimar, U., & Vasanelli, L. (2001). Analysis of CO and CH4 gas mixtures by using a micromachined sensor array. Sensors and Actuators B: Chemical, 78(1), 40–48. https://doi.org/10.1016/S0925-4005(01)00789-4

CARB. (2017, July 25). Sites Measuring Total Hydrocarbons. Retrieved December 20, 2018, from Quality Assurance Air Monitoring Site Information website: https://www.arb.ca.gov/qaweb/sitelist_generator.php

Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., & Tang, Y. (2008). Predicting air quality: Improvements through advanced methods to integrate models and measurements. Journal of Computational Physics, 227(7), 3540–3571. https://doi.org/10.1016/j.jcp.2007.02.024

Carotta, M. C., Martinelli, G., Crema, L., Malagù, C., Merli, M., Ghiotti, G., & Traversa, E. (2001). Nanostructured thick-film gas sensors for atmospheric pollutant monitoring: quantitative analysis on field tests. Sensors and Actuators B: Chemical, 76(1), 336–342. https://doi.org/10.1016/S0925-4005(01)00620-7

Casey, J. G., Collier-Oxandale, A., & Hannigan, M. (2019a). Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. Sensors and Actuators B: Chemical, 283, 504–514. https://doi.org/10.1016/j.snb.2018.12.049

Casey, J. G., Collier-Oxandale, A., & Hannigan, M. P. (2019b). Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. Sensors and Actuators B: Chemical, 283, 504–514. https://doi.org/10.1016/j.snb.2018.12.049

Casey, J. G., & Hannigan, M. P. (2018). Testing the performance of field calibration techniques for low-cost gas sensors in new deployment locations: across a county line and across Colorado. Atmospheric Measurement Techniques Discussions, 1–31. https://doi.org/10.5194/amt-2018-81

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., … Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environment International, 99(Supplement C), 293–302. https://doi.org/10.1016/j.envint.2016.12.007

Castell, N., Kobernus, M., Liu, H.-Y., Schneider, P., Lahoz, W., Berre, A. J., & Noll, J. (2015). Mobile technologies and services for environmental monitoring: The Citi-Sense-MOB approach. Urban Climate, 14(Part 3), 370–382. https://doi.org/10.1016/j.uclim.2014.08.002

Cavaliere, A., Carotenuto, F., Di Gennaro, F., Gioli, B., Gualtieri, G., Martelli, F., … Zaldei, A. (2018). Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of PM2.5 and PM10 Sensors. Sensors, 18(9), 2843. https://doi.org/10.3390/s18092843

Cavicchi, R. E., Suehle, J. S., Kreider, K. G., Gaitan, M., & Chaparala, P. (1996). Optimized temperature-pulse sequences for the enhancement of chemically specific response patterns from micro-hotplate gas sensors. Sensors and Actuators B: Chemical, 33(1), 142–146. https://doi.org/10.1016/0925-4005(96)01821-7

Chang, Y., Tang, N., Qu, H., Liu, J., Zhang, D., Zhang, H., … Duan, X. (2016). Detection of Volatile Organic Compounds by Self-assembled Monolayer Coated Sensor Array with Concentration-independent Fingerprints. Scientific Reports, 6, srep23970. https://doi.org/10.1038/srep23970

Chen, C., Driggs Campbell, K., Negi, I., Iglesias, R. A., Owens, P., Tao, N., … Forzani, E. S. (2012). A new sensor for the assessment of personal exposure to volatile organic compounds. Atmospheric Environment, 54(Supplement C), 679–687. https://doi.org/10.1016/j.atmosenv.2012.01.048

Chen, Z., Zheng, Y., Chen, K., Li, H., & Jian, J. (2017). Concentration Estimator of Mixed VOC Gases Using Sensor Array With Neural Networks and Decision Tree Learning. IEEE Sensors Journal, 17(6), 1884–1892. https://doi.org/10.1109/JSEN.2017.2653400

Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., … Hannigan, M. (2017). Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). Sensors, 17(11), 2478. https://doi.org/10.3390/s17112478

Clemitshaw, K. (2004). A Review of Instrumentation and Measurement Techniques for Ground-Based and Airborne Field Studies of Gas-Phase Tropospheric Chemistry. Critical Reviews in Environmental Science and Technology, 34(1), 1–108. https://doi.org/10.1080/10643380490265117

Coggon, M. M., McDonald, B. C., Vlasenko, A., Veres, P. R., Bernard, F., Koss, A. R., … de Gouw, J. A. (2018). Diurnal Variability and Emission Pattern of Decamethylcyclopentasiloxane (D5) from the Application of Personal Care Products in Two North American Cities. Environmental Science & Technology, 52(10), 5610–5618. https://doi.org/10.1021/acs.est.8b00506

Collier, A. M., Hannigan, M., Piedrahita, R., Casey, J. G., Johnston, J., & Chiang, S. (2016). Quantifying a Total Non-Methane Hydrocarbon Signal using Low-Cost VOC Sensors in an Effort to Help Communities Learn More About their Air Quality. AGU Fall Meeting Abstracts, 21. Retrieved from http://adsabs.harvard.edu/abs/2016AGUFM.A21G0154C

Collier-Oxandale, A., Hannigan, M. P., Casey, J. G., Piedrahita, R., Ortega, J., Halliday, H., & Johnston, J. (2018). Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. Atmos. Meas. Tech. Discuss., 2018, 1–35. https://doi.org/10.5194/amt-2017-421

Collier-Oxandale, A. M., Thorson, J., Halliday, H., Milford, J., & Hannigan, M. (2018). Understanding the ability of low-cost MOx sensors to quantify ambient VOCs. Atmospheric Measurement Techniques Discussions, 1–29. https://doi.org/10.5194/amt-2018-304

Collier-Oxandale, Ashley, Coffey, E., Thorson, J., Johnston, J., & Hannigan, M. (2018). Comparing Building and Neighborhood-Scale Variability of CO2 and O3 to Inform Deployment Considerations for Low-Cost Sensor System Use. Sensors (Basel, Switzerland), 18(5). https://doi.org/10.3390/s18051349

Craven, M. A., Gardner, J. W., & Bartlett, P. N. (1996). Electronic noses — development and future prospects. TrAC Trends in Analytical Chemistry, 15(9), 486–493. https://doi.org/10.1016/S0165-9936(96)00061-1

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., … Jayne, J. T. (2017). Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmos. Meas. Tech., 10(9), 3575–3588. https://doi.org/10.5194/amt-10-3575-2017

Currie, J., Greenstone, M., & Meckel, K. (2017). Hydraulic fracturing and infant health: New evidence from Pennsylvania. Science Advances, 3(12), e1603021. https://doi.org/10.1126/sciadv.1603021

de Gouw, J. A., Gilman, J. B., Kim, S.-W., Alvarez, S. L., Dusanter, S., Graus, M., … Young, C. J. (2018). Chemistry of Volatile Organic Compounds in the Los Angeles Basin: Formation of Oxygenated Compounds and Determination of Emission Ratios. Journal of Geophysical Research: Atmospheres, 123(4), 2298–2319. https://doi.org/10.1002/2017JD027976

De Vito, S., Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., & Di Francia, G. (2018). Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches. Sensors and Actuators B: Chemical, 255, 1191–1210. https://doi.org/10.1016/j.snb.2017.07.155

De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators B: Chemical, 129(2), 750–757. https://doi.org/10.1016/j.snb.2007.09.060

De Vito, Saverio, Castaldo, A., Loffredo, F., Massera, E., Polichetti, T., Nasti, I., … Di Francia, G. (2007). Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures. Sensors and Actuators B: Chemical, 124(2), 309–316. https://doi.org/10.1016/j.snb.2006.12.039

De Vito, Saverio, Piga, M., Martinotto, L., & Di Francia, G. (2009). CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. Sensors and Actuators B: Chemical, 143(1), 182–191. https://doi.org/10.1016/j.snb.2009.08.041

Dedering, U. (2010). Location map of the USA (without Hawaii and Alaska). Retrieved from https://commons.wikimedia.org/wiki/File:Usa_edcp_location_map.svg

Delgado-Saborit, J. M., Aquilina, N. J., Meddings, C., Baker, S., & Harrison, R. M. (2011). Relationship of personal exposure to volatile organic compounds to home, work and fixed site outdoor concentrations. Science of The Total Environment, 409(3), 478–488. https://doi.org/10.1016/j.scitotenv.2010.10.014

Deng, Y., Chen, C., Xian, X., Tsow, F., Verma, G., McConnell, R., … Forzani, E. S. (2016). A Novel Wireless Wearable Volatile Organic Compound (VOC) Monitoring Device with Disposable Sensors. Sensors, 16(12), 2060. https://doi.org/10.3390/s16122060

Ding, J., McAvoy, T. J., Cavicchi, R. E., & Semancik, S. (2001). Surface state trapping models for SnO2-based microhotplate sensors. Sensors and Actuators B: Chemical, 77(3), 597–613. https://doi.org/10.1016/S0925-4005(01)00765-1

Directive 2008/50/EC. , (2008).

Ditto, J. C., Barnes, E. B., Khare, P., Takeuchi, M., Joo, T., Bui, A. A. T., … Gentner, D. R. (2018). An omnipresent diversity and variability in the chemical composition of atmospheric functionalized organic aerosol. Communications Chemistry, 1(1), 75. https://doi.org/10.1038/s42004-018-0074-3

Dominguez-Pumar, M., Kowalski, L., Calavia, R., & Llobet, E. (2016). Smart control of chemical gas sensors for the reduction of their time response. Sensors and Actuators B: Chemical, 229, 1–6. https://doi.org/10.1016/j.snb.2016.01.081

Dorsey, K. L., & Pisano, A. P. (2016). Stability and Control of a Metal Oxide Gas Sensor in Simulated Wind. IEEE Sensors Journal, 16(3), 699–705. https://doi.org/10.1109/JSEN.2015.2485939

Eisele, A. P., Mukerjee, S., Smith, L. A., Thoma, E. D., Whitaker, D. A., Oliver, K. D., … Stallings, C. (2016). Volatile organic compounds at two oil and natural gas production well pads in Colorado and Texas using passive samplers. Journal of the Air & Waste Management Association, 66(4), 412–419. https://doi.org/10.1080/10962247.2016.1141808

ESA. (2017). MERLIN (Methane Remote Sensing Lidar Mission) Minisatellite. Retrieved from eoPortal website: https://directory.eoportal.org/web/eoportal/satellite-missions/content/-/article/merlin

Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R. L., & Popoola, O. (2016). Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. Sensors and Actuators B: Chemical, 231(Supplement C), 701–713. https://doi.org/10.1016/j.snb.2016.03.038

Eugster, W., & Kling, G. W. (2012). Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies. Atmos. Meas. Tech., 5(8), 1925–1934. https://doi.org/10.5194/amt-5-1925-2012

European Environmental Agency. (2017, November 20). Contribution of the transport sector to total emissions of the main air pollutants [Data Visualization]. Retrieved November 18, 2018, from European Environment Agency website: https://www.eea.europa.eu/data-and-maps/daviz/contribution-of-the-transport-sector-4

Faiz, A. (1996). Air pollution from motor vehicles : standards and technologies for controlling emissions (No. 16190; p. 1). Retrieved from The World Bank website: http://documents.worldbank.org/curated/en/606461468739185298/Air-pollution-from-motor-vehicles-standards-and-technologies-for-controlling-emissions

Feinberg, S., Williams, R., Hagler, G. S. W., Rickard, J., Brown, R., Garver, D., … Garvey, S. (2018). Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado. Atmospheric Measurement Techniques, 11(8), 4605–4615. https://doi.org/10.5194/amt-11-4605-2018

Fernandez, L., Guney, S., Gutierrez-Galvez, A., & Marco, S. (2016). Calibration transfer in temperature modulated gas sensor arrays. Sensors and Actuators B: Chemical, 231, 276–284. https://doi.org/10.1016/j.snb.2016.02.131

Fernandez, Luis, Yan, J., Fonollosa, J., Burgués, J., Gutierrez, A., & Marco, S. (2018). A Practical Method to Estimate the Resolving Power of a Chemical Sensor Array: Application to Feature Selection. Frontiers in Chemistry, 6. https://doi.org/10.3389/fchem.2018.00209

Feudale, R. N., Woody, N. A., Tan, H., Myles, A. J., Brown, S. D., & Ferré, J. (2002). Transfer of multivariate calibration models: a review. Chemometrics and Intelligent Laboratory Systems, 64(2), 181–192. https://doi.org/10.1016/S0169-7439(02)00085-0

Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D. M., … Bartonova, A. (2017). An evaluation tool kit of air quality micro-sensing units. Science of The Total Environment, 575, 639–648. https://doi.org/10.1016/j.scitotenv.2016.09.061

Fonollosa, J., Fernández, L., Gutiérrez-Gálvez, A., Huerta, R., & Marco, S. (2016a). Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. Sensors and Actuators B: Chemical, 236, 1044–1053. https://doi.org/10.1016/j.snb.2016.05.089

Fonollosa, J., Fernández, L., Gutiérrez-Gálvez, A., Huerta, R., & Marco, S. (2016b). Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. Sensors and Actuators B: Chemical, 236, 1044–1053. https://doi.org/10.1016/j.snb.2016.05.089

Fonollosa, J., Neftci, E., Huerta, R., & Marco, S. (2015). Evaluation of calibration transfer strategies between Metal Oxide gas sensor arrays. Procedia Engineering, 120, 261–264. https://doi.org/10.1016/j.proeng.2015.08.601

Fonollosa, Jordi, Rodríguez-Luján, I., & Huerta, R. (2015). Chemical gas sensor array dataset. Data in Brief, 3, 85–89. https://doi.org/10.1016/j.dib.2015.01.003

Fonollosa, Jordi, Sheik, S., Huerta, R., & Marco, S. (2015). Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. Sensors and Actuators B: Chemical, 215, 618–629. https://doi.org/10.1016/j.snb.2015.03.028

Fort, A., Mugnaini, M., Pasquini, I., Rocchi, S., & Vignoli, V. (2011). Modeling of the influence of H2O on metal oxide sensor responses to CO. Sensors and Actuators B: Chemical, 159(1), 82–91. https://doi.org/10.1016/j.snb.2011.06.052

Foster-Wittig, T. (2015). Mobile Sensors: Assessment of Fugitive Methane Emissions from Near and Far-Field Sources (Ph.D., Duke University). Retrieved from https://search.proquest.com/docview/1677412554/abstract/449B71F8EEA440D3PQ/1

Foster-Wittig, T. A., Thoma, E. D., & Albertson, J. D. (2015). Estimation of point source fugitive emission rates from a single sensor time series: A conditionally-sampled Gaussian plume reconstruction. Atmospheric Environment, 115, 101–109. https://doi.org/10.1016/j.atmosenv.2015.05.042

Gaiardo, A., Fabbri, B., Guidi, V., Bellutti, P., Giberti, A., Gherardi, S., … Zonta, G. (2016). Metal Sulfides as Sensing Materials for Chemoresistive Gas Sensors. Sensors, 16(3), 296. https://doi.org/10.3390/s16030296

Gallagher, M. E., Down, A., Ackley, R. C., Zhao, K., Phillips, N., & Jackson, R. B. (2015). Natural Gas Pipeline Replacement Programs Reduce Methane Leaks and Improve Consumer Safety. Environmental Science & Technology Letters, 2(10), 286–291. https://doi.org/10.1021/acs.estlett.5b00213

Garcia, C. AB-617 Nonvehicular air pollution: criteria air pollutants and toxic air contaminants. , Chapter 136 California Health and Safety Code § 40920.6, 42400, 42402, 39607.1, 40920.8, 42411, 42705.5, and 44391.2 (2017).

Gebicki, J., & Dymerski, T. (2016). Chapter 11 - Application of Chemical Sensors and Sensor Matrixes to Air Quality Evaluation. In M. de la Guardia & S. Armenta (Eds.), Comprehensive Analytical Chemistry (pp. 267–294). https://doi.org/10.1016/bs.coac.2016.02.007

George, D. L., Poerner, N. W., Ridens, B. L., & Thorson, J. L. (2014). Compressed Natural Gas Vehicle Fuel Survey. Retrieved from https://trid.trb.org/view/1318095

Gilman, J. B., Lerner, B. M., Kuster, W. C., Goldan, P. D., Warneke, C., Veres, P. R., … Yokelson, R. J. (2015). Biomass burning emissions and potential air quality impacts of volatile organic compounds and other trace gases from fuels common in the US. Atmospheric Chemistry and Physics, 15(24), 13915–13938. https://doi.org/10.5194/acp-15-13915-2015

Goldan, P. D., Kuster, W. C., Williams, E., Murphy, P. C., Fehsenfeld, F. C., & Meagher, J. (2004). Nonmethane hydrocarbon and oxy hydrocarbon measurements during the 2002 New England Air Quality Study. Journal of Geophysical Research: Atmospheres, 109(D21). https://doi.org/10.1029/2003JD004455

Gosangi, R., & Gutierrez-Osuna, R. (2013). Active temperature modulation of metal-oxide sensors for quantitative analysis of gas mixtures. Sensors and Actuators B: Chemical, 185, 201–210. https://doi.org/10.1016/j.snb.2013.04.056

Goyal, A., Small, M. J., von Stackelberg, K., Burmistrov, D., & Jones, N. (2005). Estimation of Fugitive Lead Emission Rates from Secondary Lead Facilities using Hierarchical Bayesian Models. Environmental Science & Technology, 39(13), 4929–4937. https://doi.org/10.1021/es035465e

Green, R. B., Hater, G. R., Thoma, E. D., DeWees, J., Rella, C. W., Crosson, E. R., … Swan, N. (2010, September 13). Methane emissions measured at two California landfills by OTM-10 and an acetylene tracer method. Presented at the Global Waste Management Symposium, San Antonio, TX. Retrieved from https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=228383&keyword=acetylene&actType=&TIMSType=+&TIMSSubTypeID=&DEID=&epaNumber=&ntisID=&archiveStatus=Both&ombCat=Any&dateBeginCreated=&dateEndCreated=&dateBeginPublishedPresented=&dateEndPublishedPresented=&dateBeginUpdated=&dateEndUpdated=&dateBeginCompleted=&dateEndCompleted=&personID=&role=Any&journalID=&publisherID=&sortBy=revisionDate&count=50&CFID=89347685&CFTOKEN=63765105

Guisasola, A., Sharma, K. R., Keller, J., & Yuan, Z. (2009). Development of a model for assessing methane formation in rising main sewers. Water Research, 43(11), 2874–2884. https://doi.org/10.1016/j.watres.2009.03.040

Gupta, S., Pebesma, E., Mateu, J., & Degbelo, A. (2018). Air Quality Monitoring Network Design Optimisation for Robust Land Use Regression Models. Sustainability, 10(5), 1442. https://doi.org/10.3390/su10051442

Haerri, H.-P., Macé, T., Waldén, J., Pascale, C., Niederhauser, B., Wirtz, K., … Waldén, T. (2017). Dilution and permeation standards for the generation of NO, NO 2 and SO 2 calibration gas mixtures. Measurement Science and Technology, 28(3), 035801. https://doi.org/10.1088/1361-6501/aa543d

Hagan, D. H., Isaacman-VanWertz, G., Franklin, J. P., Wallace, L. M. M., Kocar, B. D., Heald, C. L., & Kroll, J. H. (2018). Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. Atmospheric Measurement Techniques; Katlenburg-Lindau, 11(1), 315–328. http://dx.doi.org/10.5194/amt-11-315-2018

Hagler, G. S. W., Williams, R., Papapostolou, V., & Polidori, A. (2018). Air Quality Sensors and Data Adjustment Algorithms: When Is It No Longer a Measurement? Environmental Science & Technology, 52(10), 5530–5531. https://doi.org/10.1021/acs.est.8b01826

Halliday, H. S., Thompson, A. M., Wisthaler, A., Blake, D. R., Hornbrook, R. S., Mikoviny, T., … Hills, A. J. (2016). Atmospheric benzene observations from oil and gas production in the Denver-Julesburg Basin in July and August 2014. Journal of Geophysical Research: Atmospheres, 121(18), 11,055-11,074. https://doi.org/10.1002/2016JD025327

Harley, R. A., Hannigan, M. P., & Cass, G. R. (1992). Respeciation of organic gas emissions and the detection of excess unburned gasoline in the atmosphere. Environmental Science & Technology, 26(12), 2395–2408. https://doi.org/10.1021/es00036a010

Heath, G., Warner, E., Steinberg, D., & Brandt, A. (2015). Estimating U.S. Methane Emissions from the Natural Gas Supply Chain. Approaches, Uncertainties, Current Estimates, and Future Studies (No. NREL/TP-6A50-62820). https://doi.org/10.2172/1226158

Heimann, I., Bright, V. B., McLeod, M. W., Mead, M. I., Popoola, O. A. M., Stewart, G. B., & Jones, R. L. (2015). Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. Atmospheric Environment, 113(Supplement C), 10–19. https://doi.org/10.1016/j.atmosenv.2015.04.057

Helwig, N., Schüler, M., Bur, C., Schütze, A., & Sauerwald, T. (2014). Gas mixing apparatus for automated gas sensor characterization. Measurement Science and Technology, 25(5), 055903. https://doi.org/10.1088/0957-0233/25/5/055903

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., … Zhang, Q. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). Geosci. Model Dev., 11(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Howard, C. J., Kumar, A., Malkina, I., Mitloehner, F., Green, P. G., Flocchini, R. G., & Kleeman, M. J. (2010). Reactive Organic Gas Emissions from Livestock Feed Contribute Significantly to Ozone Production in Central California. Environmental Science & Technology, 44(7), 2309–2314. https://doi.org/10.1021/es902864u

Huerta, R., Mosqueiro, T., Fonollosa, J., Rulkov, N. F., & Rodriguez-Lujan, I. (2016). Online decorrelation of humidity and temperature in chemical sensors for continuous monitoring. Chemometrics and Intelligent Laboratory Systems, 157, 169–176. https://doi.org/10.1016/j.chemolab.2016.07.004

Humayun, M. T. (ORCID:0000000214847283), Divan, R., Stan, L., Rosenmann, D., Gosztola, D., Gundel, L., … Paprotny, I. (2016). Ubiquitous Low-Cost Functionalized Multi-Walled Carbon Nanotube Sensors for Distributed Methane Leak Detection. IEEE Sensors Journal, 16(24). https://doi.org/10.1109/JSEN.2016.2581832

Ionescu, R., & Llobet, E. (2002). Wavelet transform-based fast feature extraction from temperature modulated semiconductor gas sensors. Sensors and Actuators B: Chemical, 81(2), 289–295. https://doi.org/10.1016/S0925-4005(01)00968-6

Isaksen, I. S. A., Berntsen, T. K., Dalsøren, S. B., Eleftheratos, K., Orsolini, Y., Rognerud, B., … Holmes, C. D. (2014). Atmospheric Ozone and Methane in a Changing Climate. Atmosphere, 5(3), 518–535. https://doi.org/10.3390/atmos5030518

Jacob, D. J. (1999). Introduction to Atmospheric Chemistry. Retrieved from http://acmg.seas.harvard.edu/people/faculty/djj/book/

Jenkins, P. L., Phillips, T. J., Mulberg, E. J., & Hui, S. P. (1992). Activity patterns of Californians: Use of and proximity to indoor pollutant sources. Atmospheric Environment. Part A. General Topics, 26(12), 2141–2148. https://doi.org/10.1016/0960-1686(92)90402-7

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., … Buckley, K. (2016). Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. Atmos. Meas. Tech., 9(11), 5281–5292. https://doi.org/10.5194/amt-9-5281-2016

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., … Worsnop, D. R. (2009). Evolution of Organic Aerosols in the Atmosphere. Science, 326(5959), 1525–1529. https://doi.org/10.1126/science.1180353

Johnson, K., & Knapp, A. (2017). Selectivity measure for arrays of non-specific sensors. Sensors and Actuators B: Chemical, 251, 1076–1088. https://doi.org/10.1016/j.snb.2017.05.182

Jurdak, R., Nafaa, A., & Barbirato, A. (2008). Large Scale Environmental Monitoring through Integration of Sensor and Mesh Networks. Sensors (Basel, Switzerland), 8(11), 7493–7517. https://doi.org/10.3390/s8117493

Kamionka, M., Breuil, P., & Pijolat, C. (2006). Calibration of a multivariate gas sensing device for atmospheric pollution measurement. Sensors and Actuators B: Chemical, 118(1), 323–327. https://doi.org/10.1016/j.snb.2006.04.058

Karelin, A., Karpov, E., Baranov, A., Mironov, S., & Karpova, E. (2017). Method for Determining the Concentration of Unknown Combustible Gas. Proceedings, 1(4), 422. https://doi.org/10.3390/proceedings1040422

Katzenstein, A. S., Doezema, L. A., Simpson, I. J., Blake, D. R., & Rowland, F. S. (2003). Extensive regional atmospheric hydrocarbon pollution in the southwestern United States. Proceedings of the National Academy of Sciences, 100(21), 11975–11979. https://doi.org/10.1073/pnas.1635258100

Keats, A., Yee, E., & Lien, F.-S. (2007). Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment, 41(3), 465–479. https://doi.org/10.1016/j.atmosenv.2006.08.044

Keeling, R. F., & Shertz, S. R. (1992). Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle. Nature, 358(6389), 723–727. https://doi.org/10.1038/358723a0

Khordakova, D. (2017). Diurnal behavior and temporal patterns of atmospheric peroxy radicals in a semi-urban area during summer 2016 (Master's, Institute of Environmental Physics). Retrieved from http://www.iup.uni-bremen.de/PEP_master_thesis/thesis_2017/Khordakova_Dina_MScThesis.pdf

Kim, J., Shusterman, A. A., Lieschke, K. J., Newman, C., & Cohen, R. C. (2017). The BErkeley Atmospheric CO2 Observation Network: Field Calibration and Evaluation of Low-cost Air Quality Sensors. Atmos. Meas. Tech. Discuss., 2017, 1–20. https://doi.org/10.5194/amt-2017-339

Klems, J. P., Pennington, M. R., Zordan, C. A., & Johnston, M. V. (2010). Ultrafine Particles Near a Roadway Intersection: Origin and Apportionment of Fast Changes in Concentration. Environmental Science & Technology, 44(20), 7903–7907. https://doi.org/10.1021/es102009e

Korotcenkov, G., & Cho, B. K. (2011). Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey). Sensors and Actuators B: Chemical, 156(2), 527–538. https://doi.org/10.1016/j.snb.2011.02.024

Korotcenkov, G., & Cho, B. K. (2017). Metal oxide composites in conductometric gas sensors: Achievements and challenges. Sensors and Actuators B: Chemical, 244, 182–210. https://doi.org/10.1016/j.snb.2016.12.117

Korotcenkov, Ghenadii. (2012). Chemical Sensors: Simulation and Modeling. Retrieved from http://library.books24x7.com.colorado.idm.oclc.org/toc.aspx?bookid=40508

Korotcenkov, Ghenadii. (2013). Handbook of Gas Sensor Materials (Ghenadii Korotcenkov, Ed.). In Vol. 1: Conventional Approaches. https://doi.org/10.1007/978-1-4614-7165-3_11

Kourtidis, K. A., Ziomas, I., Zerefos, C., Kosmidis, E., Symeonidis, P., Christophilopoulos, E., … Mploutsos, A. (2002). Benzene, toluene, ozone, NO2 and SO2 measurements in an urban street canyon in Thessaloniki, Greece. Atmospheric Environment, 36(34), 5355–5364. https://doi.org/10.1016/S1352-2310(02)00580-0

Lai, A. M., Carter, E., Shan, M., Ni, K., Clark, S., Ezzati, M., … Schauer, J. J. (2019). Chemical composition and source apportionment of ambient, household, and personal exposures to PM2.5 in communities using biomass stoves in rural

China. Science of The Total Environment, 646, 309–319. https://doi.org/10.1016/j.scitotenv.2018.07.322

Lee, A. P., & Reedy, B. J. (1999). Temperature modulation in semiconductor gas sensing. Sensors and Actuators B: Chemical, 60(1), 35–42. https://doi.org/10.1016/S0925-4005(99)00241-5

Leidinger, M., Sauerwald, T., Alépée, C., & Schütze, A. (2016). Miniaturized Integrated Gas Sensor Systems Combining Metal Oxide Gas Sensors and Pre-concentrators. Procedia Engineering, 168(Supplement C), 293–296. https://doi.org/10.1016/j.proeng.2016.11.199

Leidinger, M., Sauerwald, T., Conrad, T., Reimringer, W., Ventura, G., & Schütze, A. (2014). Selective Detection of Hazardous Indoor VOCs Using Metal Oxide Gas Sensors. Procedia Engineering, 87, 1449–1452. https://doi.org/10.1016/j.proeng.2014.11.722

Leidinger, Martin, Rieger, M., Sauerwald, T., Alépée, C., & Schütze, A. (2016). Integrated pre-concentrator gas sensor microsystem for ppb level benzene detection. Sensors and Actuators B: Chemical, 236(Supplement C), 988–996. https://doi.org/10.1016/j.snb.2016.04.064

Leidinger, Martin, Schultealbert, C., Neu, J., Schütze, A., & Sauerwald, T. (2018). Characterization and calibration of gas sensor systems at ppb level—a versatile test gas generation system. Measurement Science and Technology, 29(1), 015901. https://doi.org/10.1088/1361-6501/aa91da

Lenschow, P., Abraham, H.-J., Kutzner, K., Lutz, M., Preuß, J.-D., & Reichenbächer, W. (2001). Some ideas about the sources of PM10. Atmospheric Environment, 35(Supplement 1), S23–S33. https://doi.org/10.1016/S1352-2310(01)00122-4

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., … White, A. (2016). Evaluating the performance of low cost chemical sensors for air pollution research. Faraday Discussions, 189(0), 85–103. https://doi.org/10.1039/C5FD00201J

Lewis, A., & Edwards, P. (2016). Validate personal air-pollution sensors. Nature News, 535(7610), 29. https://doi.org/10.1038/535029a

Lewis, A., Peltier, W. R., & von Schneidemesser, E. (2018). Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications [Monograph]. Retrieved from World Meteorological Organization website: http://eprints.whiterose.ac.uk/135994/

Lin, C.-H., & Chang, L.-F. W. (2002). Relative source contribution analysis using an air trajectory statistical approach. Journal of Geophysical Research: Atmospheres, 107(D21), 4583. https://doi.org/10.1029/2001JD001301

Linquist, B., Groenigen, K. J. van, Adviento-Borbe, M. A., Pittelkow, C., & Kessel, C. van. (2012). An agronomic assessment of greenhouse gas emissions from major cereal crops. Global Change Biology, 18(1), 194–209. https://doi.org/10.1111/j.1365-2486.2011.02502.x

Magna, G., Mosciano, F., Martinelli, E., & Di Natale, C. (2018). Unsupervised On-Line Selection of Training Features for a robust classification with drifting and faulty gas sensors. Sensors and Actuators B: Chemical, 258, 1242–1251. https://doi.org/10.1016/j.snb.2017.12.005

Marco, S., & Gutierrez-Galvez, A. (2012). Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review. IEEE Sensors Journal, 12(11), 3189–3214. https://doi.org/10.1109/JSEN.2012.2192920

Marle, M. J. E. van, Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., … Werf, G. R. van der. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). Geoscientific Model Development, 10(9), 3329–3357. https://doi.org/10.5194/gmd-10-3329-2017

Martenies, S. E., Milando, C. W., Williams, G. O., & Batterman, S. A. (2017). Disease and Health Inequalities Attributable to Air Pollutant Exposure in Detroit, Michigan. International Journal of Environmental Research and Public Health, 14(10), 1243. https://doi.org/10.3390/ijerph14101243

Martinelli, E., Magna, G., De Vito, S., Di Fuccio, R., Di Francia, G., Vergara, A., & Di Natale, C. (2013). An adaptive classification model based on the Artificial Immune System for chemical sensor drift mitigation. Sensors and Actuators B: Chemical, 177, 1017–1026. https://doi.org/10.1016/j.snb.2012.11.107

Masiol, M., Squizzato, S., Chalupa, D., Rich, D. Q., & Hopke, P. K. (2018). Spatial-temporal variations of summertime ozone concentrations across a metropolitan area using a network of low-cost monitors to develop 24 hourly land-use regression models. Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2018.11.111

Masson, N., Piedrahita, R., & Hannigan, M. (2015). Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring. Sensors and Actuators B: Chemical, 208, 339–345. https://doi.org/10.1016/j.snb.2014.11.032

Masson, Nicholas, Piedrahita, R., & Hannigan, M. (2015). Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality. Sensors, 15(10), 27283–27302. https://doi.org/10.3390/s151027283

McDonald, B. C., Gouw, J. A. de, Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., … Trainer, M. (2018a). Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Science, 359(6377), 760–764. https://doi.org/10.1126/science.aaq0524

McDonald, B. C., Gouw, J. A. de, Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., … Trainer, M. (2018b). Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Science, 359(6377), 760–764. https://doi.org/10.1126/science.aaq0524

McDuffie, E. E., Edwards, P. M., Gilman, J. B., Lerner, B. M., Dubé, W. P., Trainer, M., … Brown, S. S. (2016). Influence of oil and gas emissions on summertime ozone in the Colorado Northern Front Range. Journal of Geophysical Research: Atmospheres, 121(14), 8712–8729. https://doi.org/10.1002/2016JD025265

McKenzie, L. M., Witter, R. Z., Newman, L. S., & Adgate, J. L. (2012). Human health risk assessment of air emissions from development of unconventional natural gas resources. Science of The Total Environment, 424, 79–87. https://doi.org/10.1016/j.scitotenv.2012.02.018

McKenzie Lisa M., Guo Ruixin, Witter Roxana Z., Savitz David A., Newman Lee S., & Adgate John L. (2014). Birth Outcomes and Maternal Residential Proximity to Natural Gas Development in Rural Colorado. Environmental Health Perspectives, 122(4), 412–417. https://doi.org/10.1289/ehp.1306722

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., … Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmospheric Environment, 70(Supplement C), 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060

Mihelcic, J. R., Baillod, C. R., Crittenden, J. C., & Rogers, T. N. (1993). Estimation of VOC Emissions from Wastewater Facilities by Volatilization and Stripping. Air & Waste, 43(1), 97–105. https://doi.org/10.1080/1073161X.1993.10467120

Mijling, B., Jiang, Q., de Jonge, D., & Bocconi, S. (2017). Practical field calibration of electrochemical NO2 sensors for urban air quality applications. Atmos. Meas. Tech. Discuss., 2017, 1–25. https://doi.org/10.5194/amt-2017-43

Mirzaei, A., Leonardi, S. G., & Neri, G. (2016). Detection of hazardous volatile organic compounds (VOCs) by metal oxide nanostructures-based gas sensors: A review. Ceramics International, 42(14), 15119–15141. https://doi.org/10.1016/j.ceramint.2016.06.145

Miskell, G., Salmond, J., & Williams, D. (2017). Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments. Science of The Total Environment, 575, 1119–1129. https://doi.org/10.1016/j.scitotenv.2016.09.177

Miskell, Georgia, Salmond, J. A., & Williams, D. E. (2018). Solution to the Problem of Calibration of Low-Cost Air Quality Measurement Sensors in Networks. ACS Sensors, 3(4), 832–843. https://doi.org/10.1021/acssensors.8b00074

Miskell, Georgia, Salmond, J., Alavi-Shoshtari, M., Bart, M., Ainslie, B., Grange, S., … Williams, D. E. (2016). Data Verification Tools for Minimizing Management Costs of Dense Air-Quality Monitoring Networks. Environmental Science & Technology, 50(2), 835–846. https://doi.org/10.1021/acs.est.5b04421

Miskell, Georgia, Salmond, J., Grange, S., Weissert, L., Henshaw, G., & Williams, D. (2017). Reliable Long-Term Data from Low-Cost Gas Sensor Networks in the Environment. Proceedings, 1(4), 400. https://doi.org/10.3390/proceedings1040400

Mitchell, L. E., Crosman, E. T., Jacques, A. A., Fasoli, B., Leclair-Marzolf, L., Horel, J., … Lin, J. C. (2018). Monitoring of greenhouse gases and pollutants across an urban area using a light-rail public transit platform. Atmospheric Environment, 187, 9–23. https://doi.org/10.1016/j.atmosenv.2018.05.044

Molino, A., Elen, B., Theunis, J., Ingarra, S., Van den Bossche, J., Reggente, M., & Loreto, V. (2012). The EveryAware SensorBox: a tool for community-based air quality monitoring. Workshop Sensing a Changing World.

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. Science of The Total Environment, 502(Supplement C), 537–547. https://doi.org/10.1016/j.scitotenv.2014.09.059

Moon, S. E., Lee, H.-K., Choi, N.-J., Lee, J., Choi, C. A., Yang, W. S., … Yoo, D.-J. (2013). Low power consumption micro C2H5OH gas sensor based on micro-heater and screen printing technique. Sensors and Actuators B: Chemical, 187(Supplement C), 598–603. https://doi.org/10.1016/j.snb.2013.05.002

Morel, P. (2018). Gramm: grammar of graphics plotting in Matlab. The Journal of Open Source Software. https://doi.org/10.21105/joss.00568

Moseley, P. T. (2017). Progress in the development of semiconducting metal oxide gas sensors: a review. Measurement Science and Technology, 28(8), 082001. https://doi.org/10.1088/1361-6501/aa7443

Mueller, M., Meyer, J., & Hueglin, C. (2017). Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich. Atmos. Meas. Tech., 10(10), 3783–3799. https://doi.org/10.5194/amt-10-3783-2017

Muezzinoglu, M. K., Vergara, A., Huerta, R., Nowotny, T., Rulkov, N., Abarbanel, H., … Rabinovich, M. (2009). Artificial Olfactory Brain for Mixture Identification. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), Advances in Neural Information Processing Systems 21 (pp. 1121–1128). Retrieved from http://papers.nips.cc/paper/3555-artificial-olfactory-brain-for-mixture-identification.pdf

NASA. (2011, October 19). A Look Back at a Decade of Fires [Feature]. Retrieved December 9, 2018, from https://www.nasa.gov/mission_pages/fires/main/modis-10.html

Nehorai, A., Porat, B., & Paldi, E. (1995). Detection and localization of vapor-emitting sources. IEEE Transactions on Signal Processing, 43(1), 243–253. https://doi.org/10.1109/78.365304

Niedojadlo, A., Becker, K. H., Kurtenbach, R., & Wiesen, P. (2007). The contribution of traffic and solvent use to the total NMVOC emission in a German city derived from measurements and CMB modelling. Atmospheric Environment, 41(33), 7108–7126. https://doi.org/10.1016/j.atmosenv.2007.04.056

Oliver, K. D., Cousett, T. A., Whitaker, D. A., Smith, L. A., Mukerjee, S., Stallings, C., … Henkle, S. (2017). Sample integrity evaluation and EPA method 325B interlaboratory comparison for select volatile organic compounds collected diffusively on Carbopack X sorbent tubes. Atmospheric Environment, 163, 99–106. https://doi.org/10.1016/j.atmosenv.2017.05.042

Oltra, C., Sala, R., Boso, À., & Asensio, S. L. (2017). Public engagement on urban air pollution: an exploratory study of two interventions. Environmental Monitoring and Assessment, 189(6), 296. https://doi.org/10.1007/s10661-017-6011-6

Papapostolou, V., Zhang, H., Feenstra, B. J., & Polidori, A. (2017). Development of an environmental chamber for evaluating the performance of low-cost air quality sensors under controlled conditions. Atmospheric Environment, 171, 82–90. https://doi.org/10.1016/j.atmosenv.2017.10.003

Pardo, A., Marco, S., & Samitier, J. (1998). Nonlinear inverse dynamic models of gas sensing systems based on chemical sensor arrays for quantitative measurements. IEEE Transactions on Instrumentation and Measurement, 47(3), 644–651. https://doi.org/10.1109/19.744316

Pardo, M., Faglia, G., Sberveglieri, G., Corte, M., Masulli, F., & Riani, M. (2000). A time delay neural network for estimation of gas concentrations in a mixture. Sensors and Actuators B: Chemical, 65(1), 267–269. https://doi.org/10.1016/S0925-4005(99)00336-6

Pashami, S. (2015). Change detection in metal oxide gas sensor signals for open sampling systems. DIVA. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-46845

Penza, M., Suriano, D., Pfister, V., Prato, M., & Cassano, G. (2017). Urban Air Quality Monitoring with Networked Low-Cost Sensor-Systems. Proceedings, 1(4), 573. https://doi.org/10.3390/proceedings1040573

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., … Tans, P. (2012). Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study. Journal of Geophysical Research: Atmospheres, 117(D4). https://doi.org/10.1029/2011JD016360

Pfister, G. G., Wiedinmyer, C., & Emmons, L. K. (2008). Impacts of the fall 2007 California wildfires on surface ozone: Integrating local observations with global model simulations. Geophysical Research Letters, 35(19). https://doi.org/10.1029/2008GL034747

Phillips, N. G., Ackley, R., Crosson, E. R., Down, A., Hutyra, L. R., Brondfield, M., … Jackson, R. B. (2013). Mapping urban pipeline leaks: Methane leaks across Boston. Environmental Pollution, 173, 1–4. https://doi.org/10.1016/j.envpol.2012.11.003

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., … Shang, L. (2014). The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmospheric Measurement Techniques; Katlenburg-Lindau, 7(10), 3325. http://dx.doi.org/10.5194/amt-7-3325-2014

Pogány, A., Balslev-Harder, D., Braban, C. F., Cassidy, N., Ebert, V., Valerio Ferracci, … Bernhard Niederhauser. (2016). A metrological approach to improve accuracy and reliability of ammonia measurements in ambient air. Measurement Science and Technology, 27(11), 115012. https://doi.org/10.1088/0957-0233/27/11/115012

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., … Jones, R. L. (2018). Use of networks of low cost air quality sensors to quantify air quality in urban settings. Atmospheric Environment, 194, 58–70. https://doi.org/10.1016/j.atmosenv.2018.09.030

Portnoff, M., Grace, R., Guzman, A., & Hibner, J. (1993). MEASUREMENT AND ANALYSIS OF ADSISTOR AND FIGARO GAS SENSORS USED FOR

UNDERGROUND STORAGE TANK LEAK DETECTION (No. EPA/600/R-92/219). Retrieved from US EPA website: https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryID=39314

Portnoff, M., Grace, R., & Hibner, J. (1995). MEASUREMENT AND ANALYSIS OF VAPOR SENSORS USED AT UNDERGROUND STORAGE TANK SITES (No. EPA/600/R-95/078 (NTIS PB96101415)). Retrieved from https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=48051&keyword=butane&actType=&TIMSType=+&TIMSSubTypeID=&DEID=&epaNumber=&ntisID=&archiveStatus=Both&ombCat=Any&dateBeginCreated=&dateEndCreated=&dateBeginPublishedPresented=&dateEndPublishedPresented=&dateBeginUpdated=&dateEndUpdated=&dateBeginCompleted=&dateEndCompleted=&personID=&role=Any&journalID=&publisherID=&sortBy=revisionDate&count=50&CFID=49446798&CFTOKEN=88101735

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., … Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. Science of The Total Environment, 607–608(Supplement C), 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266

Rella, C. W., Tsai, T. R., Botkin, C. G., Crosson, E. R., & Steele, D. (2015). Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique. Environmental Science & Technology, 49(7), 4742–4748. https://doi.org/10.1021/acs.est.5b00099

Robinson, J. A., Kocman, D., Horvat, M., & Bartonova, A. (2018). End-User Feedback on a Low-Cost Portable Air Quality Sensor System—Are We There Yet? Sensors, 18(11), 3768. https://doi.org/10.3390/s18113768

Rodriguez-Lujan, I., Fonollosa, J., Vergara, A., Homer, M., & Huerta, R. (2014). On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. Chemometrics and Intelligent Laboratory Systems, 130, 123–134. https://doi.org/10.1016/j.chemolab.2013.10.012

Romain, A. C., & Nicolas, J. (2010). Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. Sensors and Actuators B: Chemical, 146(2), 502–506. https://doi.org/10.1016/j.snb.2009.12.027

Ruckstuhl, A. F., Henne, S., Reimann, S., Steinbacher, M., M. K. Vollmer, O'Doherty, S., … Hueglin, C. (2012). Robust extraction of baseline signal of atmospheric trace species using local regression. Atmos. Meas. Tech., 5(11), 2613–2624. https://doi.org/10.5194/amt-5-2613-2012

Ryerson, T. B., Aikin, K. C., Angevine, W. M., Atlas, E. L., Blake, D. R., Brock, C. A., … Watts, L. A. (2011). Atmospheric emissions from the Deepwater Horizon spill constrain air-water partitioning, hydrocarbon fate, and leak rate. Geophysical Research Letters, 38(7). https://doi.org/10.1029/2011GL046726

Sabaliauskas, K., Jeong, C.-H., Yao, X., & Evans, G. J. (2014). The application of wavelet decomposition to quantify the local and regional sources of ultrafine particles in cities. Atmospheric Environment, 95, 249–257. https://doi.org/10.1016/j.atmosenv.2014.05.035

Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D. K., & Hannigan, M. (2018). Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors. Atmospheric Measurement Techniques, 11(3), 1777–1792. https://doi.org/10.5194/amt-11-1777-2018

Sahagun, L. (2013, November 8). EPA officers sickened by fumes at South L.A. oil field. Los Angeles Times. Retrieved from https://www.latimes.com/local/la-me-1109-fumes-20131109-story.html

Sauerwald, T., Baur, T., Leidinger, M., Reimringer, W., Spinelle, L., Gerboles, M., … Schütze, A. (2018). Highly sensitive benzene detection with metal oxide semiconductor gas sensors – an inter-laboratory comparison. Journal of Sensors and Sensor Systems, 7(1), 235–243. https://doi.org/10.5194/jsss-7-235-2018

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., … Zhu, Q. (2016). The global methane budget 2000–2012. Earth System Science Data, 8(2), 697–751. https://doi.org/10.5194/essd-8-697-2016

Scheepers, P. T. J., Van Wel, L., Beckmann, G., & Anzion, R. B. M. (2017). Chemical Characterization of the Indoor Air Quality of a University Hospital: Penetration of Outdoor Air Pollutants. International Journal of Environmental Research and Public Health, 14(5). https://doi.org/10.3390/ijerph14050497

Schmuker, M., Bahr, V., & Huerta, R. (2016). Exploiting plume structure to decode gas source distance using metal-oxide gas sensors. Sensors and Actuators B: Chemical, 235, 636–646. https://doi.org/10.1016/j.snb.2016.05.098

Schneider, P., Castell, N., Dauge, F. R., Vogt, M., Lahoz, W. A., & Bartonova, A. (2018). A Network of Low-Cost Air Quality Sensors and Its Use for Mapping Urban Air Quality. In Earth Systems Data and Models. Mobile Information Systems Leveraging Volunteered Geographic Information for Earth Observation (pp. 93–110). https://doi.org/10.1007/978-3-319-70878-2_5

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environment International, 106, 234–247. https://doi.org/10.1016/j.envint.2017.05.005

Schüler, M., Fricke, T., Sauerwald, T., & Schütze, A. (2015). E8.4 - Detecting poisoning of metal oxide gas sensors at an early stage by temperature cycled operation. Proceedings SENSOR 2015, 735–739. http://dx.doi.org/10.5162/sensor2015/E8.4

Schüler, M., Sauerwald, T., & Schütze, A. (2014). Metal oxide semiconductor gas sensor self-test using Fourier-based impedance spectroscopy. Journal of Sensors and Sensor Systems, 3(2), 213–221. https://doi.org/10.5194/jsss-3-213-2014

Schultealbert, C., Baur, T., Schütze, A., Böttcher, S., & Sauerwald, T. (2017). A novel approach towards calibrated measurement of trace gases using metal oxide semiconductor sensors. Sensors and Actuators B: Chemical, 239(Supplement C), 390–396. https://doi.org/10.1016/j.snb.2016.08.002

Schultealbert, C., Baur, T., Schütze, A., & Sauerwald, T. (2018). Facile Quantification and Identification Techniques for Reducing Gases over a Wide Concentration Range Using a MOS Sensor in Temperature-Cycled Operation. Sensors, 18(3), 744. https://doi.org/10.3390/s18030744

Schütze, A., Baur, T., Leidinger, M., Reimringer, W., Jung, R., Conrad, T., & Sauerwald, T. (2017). Highly Sensitive and Selective VOC Sensor Systems Based on Semiconductor Gas Sensors: How to? Environments, 4(1), 20. https://doi.org/10.3390/environments4010020

Schuyler, T. J., & Guzman, M. I. (2017). Unmanned Aerial Systems for Monitoring Trace Tropospheric Gases. Atmosphere, 8(10), 206. https://doi.org/10.3390/atmos8100206

Sears, W. M., Colbow, K., & Consadori, F. (1989). General characteristics of thermally cycled tin oxide gas sensors. Semiconductor Science and Technology, 4(5), 351. https://doi.org/10.1088/0268-1242/4/5/004

Seinfeld, J. H., & Pandis, S. N. (2016). Atmospheric Chemistry and Physics: From Air Pollutaion to Climate Change (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Shai Shalev-Shwartz, & Shai Ben-David. (2014). Understanding Machine Learning: From Theory to Algorithms. Retrieved from http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/

Shamasunder, B., Collier-Oxandale, A., Blickley, J., Sadd, J., Chan, M., Navarro, S., … Wong, N. J. (2018). Community-Based Health and Exposure Study around Urban Oil Developments in South Los Angeles. International Journal of Environmental Research and Public Health, 15(1), 138. https://doi.org/10.3390/ijerph15010138

Siebenaler, S. P., Janka, A. M., Lyon, D., Edlebeck, J. P., & Nowlan, A. E. (2016). Methane Detectors Challenge: Low-Cost Continuous Emissions Monitoring. V003T04A013. https://doi.org/10.1115/IPC2016-64670

Skiba, Y. N. (2003). On a method of detecting the industrial plants which violate prescribed emission rates. Ecological Modelling, 159(2), 125–132. https://doi.org/10.1016/S0304-3800(02)00279-X

Smith, K. R., Edwards, P. M., Evans, M. J., Lee, J. D., Shaw, M. D., Squires, F., … Lewis, A. C. (2017). Clustering approaches to improve the performance of low cost air pollution sensors. Faraday Discussions, 200(0), 621–637. https://doi.org/10.1039/C7FD00020K

Sohn, J. H., Atzeni, M., Zeller, L., & Pioggia, G. (2008). Characterisation of humidity dependence of a metal oxide semiconductor sensor array using partial least squares. Sensors and Actuators B: Chemical, 131(1), 230–235. https://doi.org/10.1016/j.snb.2007.11.009

Solórzano, A., Rodríguez-Pérez, R., Padilla, M., Graunke, T., Fernandez, L., Marco, S., & Fonollosa, J. (2018). Multi-unit calibration rejects inherent device variability of chemical sensor arrays. Sensors and Actuators B: Chemical, 265, 142–154. https://doi.org/10.1016/j.snb.2018.02.188

Song, K., Wang, Q., Liu, Q., Zhang, H., & Cheng, Y. (2011). A Wireless Electronic Nose System Using a Fe2O3 Gas Sensing Array and Least Squares Support Vector Regression. Sensors (Basel, Switzerland), 11(1), 485–505. https://doi.org/10.3390/s110100485

Spinelle, L., Gerboles, M., Kok, G., & Sauerwald, T. (2015). 02 - Sensitivity of VOC Sensors for Air Quality Monitoring within the EURAMET Key-VOC project. Proceedings, 6–9. http://dx.doi.org/10.5162/4EuNetAir2015/02

Spinelle, Laurent, Aleixandre, M., & Gerboles, M. (2013). Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution [EUR - Scientific and Technical Research Reports]. https://doi.org/10.2788/9916 (online)

Spinelle, Laurent, Gerboles, M., & Aleixandre, M. (2014). Report of laboratory and in-situ validation of micro-sensor for monitoring ambient : O12: CairClipO3/NO2 of CAIRPOL (F). (No. JRC86479). Retrieved from Institute for Environment and Sustainability website: https://publications.europa.eu/en/publication-detail/-/publication/3b9ffd1f-d272-4e40-a6e7-2ab8ae5de7f9/language-en

Spinelle, Laurent, Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017a). Performance Evaluation of Low-Cost BTEX Sensors and Devices within the

EURAMET Key-VOCs Project. Proceedings, 1(4), 425. https://doi.org/10.3390/proceedings1040425

Spinelle, Laurent, Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017b). Review of Portable and Low-Cost Sensors for the Ambient Air Monitoring of Benzene and Other Volatile Organic Compounds. Sensors, 17(7), 1520. https://doi.org/10.3390/s17071520

Spinelle, Laurent, Gerboles, M., Kok, G., & Sauerwald, T. (2015). Review of low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds [EUR - Scientific and Technical Research Reports]. https://doi.org/10.2788/05768

Spinelle, Laurent, Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. Sensors and Actuators B: Chemical, 215(Supplement C), 249–257. https://doi.org/10.1016/j.snb.2015.03.031

Spinelle, Laurent, Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2. Sensors and Actuators B: Chemical, 238(Supplement C), 706–715. https://doi.org/10.1016/j.snb.2016.07.036

Suárez, J. I., Arroyo, P., Lozano, J., Herrero, J. L., & Padilla, M. (2018). Bluetooth gas sensing module combined with smartphones for air quality monitoring. Chemosphere, 205, 618–626. https://doi.org/10.1016/j.chemosphere.2018.04.154

Subramanian, R., Ellis, A., Torres-Delgado, E., Tanzer, R., Malings, C., Rivera, F., … Mayol-Bracero, O. L. (2018). Air Quality in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower Cost Air Quality Monitors. ACS Earth and Space Chemistry. https://doi.org/10.1021/acsearthspacechem.8b00079

Suzuki, K., & Takada, T. (1995). Highly sensitive odour sensors using various SnO2 thick films. Sensors and Actuators B: Chemical, 25(1), 773–776. https://doi.org/10.1016/0925-4005(95)85171-2

Szczurek, A., Szecówka, P. M., & Licznerski, B. W. (1999). Application of sensor array and neural networks for quantification of organic solvent vapours in air. Sensors and Actuators B: Chemical, 58(1), 427–432. https://doi.org/10.1016/S0925-4005(99)00105-7

Szulczyński, B., & Gębicki, J. (2017). Currently Commercially Available Chemical Sensors Employed for Detection of Volatile Organic Compounds in Outdoor and Indoor Air. Environments, 4(1), 21. https://doi.org/10.3390/environments4010021

Tan, Y., Lipsky, E. M., Saleh, R., Robinson, A. L., & Presto, A. A. (2014). Characterizing the Spatial Variation of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA. Environmental Science & Technology, 48(24), 14186–14194. https://doi.org/10.1021/es5034074

Tangirala, V. K. K., Gómez-Pozos, H., Rodríguez-Lugo, V., & Olvera, M. D. L. L. (2017). A Study of the CO Sensing Responses of Cu-, Pt- and Pd-Activated SnO2 Sensors: Effect of Precipitation Agents, Dopants and Doping Methods. Sensors (Basel, Switzerland), 17(5). https://doi.org/10.3390/s17051011

Thoma, E. D., Miller, M. C., Chung, K. C., Parsons, N. L., & Shine, B. C. (2011). Facility Fence-Line Monitoring Using Passive Samplers. Journal of the Air & Waste Management Association, 61(8), 834–842. https://doi.org/10.3155/1047-3289.61.8.834

Tomchenko, A. A., Harmer, G. P., Marquis, B. T., & Allen, J. W. (2003). Semiconducting metal oxide sensor array for the selective detection of combustion gases. Sensors and Actuators B: Chemical, 93(1), 126–134. https://doi.org/10.1016/S0925-4005(03)00240-5

Tsujita, W., Yoshino, A., Ishida, H., & Moriizumi, T. (2005). Gas sensor network for air-pollution monitoring. Sensors and Actuators B: Chemical, 110(2), 304–311. https://doi.org/10.1016/j.snb.2005.02.008

US DOT. (2009, November 23). Estimated National Average Vehicle Emissions Rates per Vehicle by Vehicle Type Using Reformulated Gasoline and Diesel | Bureau of Transportation Statistics. Retrieved May 17, 2018, from https://www.bts.gov/content/estimated-national-average-vehicle-emissions-rates-vehicle-vehicle-type-using-reformulated

US EIA. (2015, April 2). Top 100 U.S. Oil and Gas Fields. Retrieved December 12, 2018, from https://www.eia.gov/naturalgas/crudeoilreserves/top100/

US EPA. (2006). Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final Report, 2006) (Reports & Assessments No. EPA/600/R-05/004aF-cF). Retrieved from U.S. Environmental Protection Agency website: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=149923

US EPA, OA. (2018). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016 (Reports and Assessments No. EPA 430-R-18-003; p. 655). Retrieved from

US EPA website: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2016

US EPA, OAR. (2014, August 19). Integrated Urban Air Toxics Strategy [Reports and Assessments]. Retrieved September 25, 2018, from US EPA website: https://www.epa.gov/urban-air-toxics/integrated-urban-air-toxics-strategy

US EPA, OAR. (2015a, August 5). Managing Air Quality - Emissions Inventories [Overviews and Factsheets]. Retrieved December 20, 2018, from US EPA website: https://www.epa.gov/air-quality-management-process/managing-air-quality-emissions-inventories

US EPA, OAR. (2015b, December 3). Hazardous Air Pollutants: Sources and Exposure [Reports and Assessments]. Retrieved November 27, 2018, from US EPA website: https://www.epa.gov/haps/hazardous-air-pollutants-sources-and-exposure

US EPA, OAR. (2016a). 2014 National Emission Inventory (NEI) Report [Policies and Guidance]. Retrieved from https://www.epa.gov/air-emissions-inventories/2014-national-emission-inventory-nei-report

US EPA, OAR. (2016b, August 10). Controlling Air Pollution from the Oil and Natural Gas Industry [Collections and Lists]. Retrieved December 10, 2018, from US EPA website: https://www.epa.gov/controlling-air-pollution-oil-and-natural-gas-industry

US EPA, OAR. (2016c, September 29). PAMS | Ambient Monitoring Technology Information Center | US EPA. Retrieved December 12, 2018, from https://www3.epa.gov/ttnamti1/pamsmain.html

US EPA, ORD. (2003, April 17). Benzene CASRN 71-43-2 | IRIS. Retrieved December 18, 2018, from https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?&substance_nmbr=276

US EPA, ORD. (2014, March 25). Basic Information about the Integrated Risk Information System [Collections and Lists]. Retrieved December 18, 2018, from US EPA website: https://www.epa.gov/iris/basic-information-about-integrated-risk-information-system

USGCRP. (2018). Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. Retrieved from U.S. Global Change Research Program website: https://nca2018.globalchange.gov/chapter/13/

van den Bossche, M., Rose, N. T., & De Wekker, S. F. J. (2017). Potential of a low-cost gas sensor for atmospheric methane monitoring. Sensors and Actuators B: Chemical, 238(Supplement C), 501–509. https://doi.org/10.1016/j.snb.2016.07.092

Vardoulakis, S., Fisher, B. E. A., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. Atmospheric Environment, 37(2), 155–182. https://doi.org/10.1016/S1352-2310(02)00857-9

Vembu, S., Vergara, A., Muezzinoglu, M. K., & Huerta, R. (2012). On time series features and kernels for machine olfaction. Sensors and Actuators B: Chemical, 174, 535–546. https://doi.org/10.1016/j.snb.2012.06.070

Vergara, A., Llobet, E., Brezmes, J., Ivanov, P., Vilanova, X., Gracia, I., … Correig, X. (2005). Optimised temperature modulation of metal oxide micro-hotplate gas sensors through multilevel pseudo random sequences. Sensors and Actuators B: Chemical, 111(Supplement C), 271–280. https://doi.org/10.1016/j.snb.2005.06.039

Vergara, Alexander, Fonollosa, J., Mahiques, J., Trincavelli, M., Rulkov, N., & Huerta, R. (2013). On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines. Sensors and Actuators B: Chemical, 185(Supplement C), 462–477. https://doi.org/10.1016/j.snb.2013.05.027

Verhulst, K. R., Karion, A., Kim, J., Salameh, P. K., Keeling, R. F., Newman, S., … Miller, C. E. (2017). Carbon dioxide and methane measurements from the Los Angeles Megacity Carbon Project – Part 1: calibration, urban enhancements, and uncertainty estimates. Atmospheric Chemistry and Physics, 17(13), 8313–8341. https://doi.org/10.5194/acp-17-8313-2017

von Schneidemesser, E., Coates, J., Denier van der Gon, H. A. C., Visschedijk, A. J. H., & Butler, T. M. (2016). Variation of the NMVOC speciation in the solvent sector and the sensitivity of modelled tropospheric ozone. Atmospheric Environment, 135, 59–72. https://doi.org/10.1016/j.atmosenv.2016.03.057

Wang, R., Moody, R. P., Koniecki, D., & Zhu, J. (2009). Low molecular weight cyclic volatile methylsiloxanes in cosmetic products sold in Canada: Implication for dermal exposure. Environment International, 35(6), 900–904. https://doi.org/10.1016/j.envint.2009.03.009

Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data. Data Mining and Knowledge Discovery, 13(3), 335–364. https://doi.org/10.1007/s10618-005-0039-x

Warneke, C., Geiger, F., Edwards, P. M., Dube, W., Pétron, G., Kofler, J., … Roberts, J. M. (2014). Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition. Atmospheric Chemistry and Physics, 14(20), 10977–10988. https://doi.org/10.5194/acp-14-10977-2014

Wen, Z., & Tian-mo, L. (2010). Gas-sensing properties of SnO2–TiO2-based sensor for volatile organic compound gas and its sensing mechanism. Physica B: Condensed Matter, 405(5), 1345–1348. https://doi.org/10.1016/j.physb.2009.11.086

Westerholm, R., & Egebäck, K. E. (1994). Exhaust emissions from light- and heavy-duty vehicles: chemical composition, impact of exhaust after treatment, and fuel parameters. Environmental Health Perspectives, 102(Suppl 4), 13–23.

Williams, D. E., Henshaw, G. S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., & Salmond, J. A. (2013). Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. Measurement Science and Technology, 24(6), 065803. https://doi.org/10.1088/0957-0233/24/6/065803

Williams, R., Long, R., Beaver, M., Kauffman, A., Zeiger, F., Heimbinder, M., … Griswold, W. (2014). Sensor Evaluation Report (No. EPA/600/R-14/143 (NTIS PB2015-100611)). Retrieved from https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=277270

Williams, Ron, Kaufman, A., & Garvey, S. (2015). Next Generation Air Monitoring (NGAM) VOC Sensor Evaluation Report (No. EPA/600/R-15/122). Retrieved from US EPA website: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=308114

Williams, Ron, Nash, D., Hagler, G., Benedict, K., MacGregor, I. C., Seay, B. A., … Dye, T. (2018). Peer Review and Supporting Literature Review of Air Sensor Technology Performance Targets (No. EPA 600/R-18/324). Retrieved from https://www.epa.gov/air-research/peer-review-and-supporting-literature-review-air-sensor-technology-performance-targets

Windischmann, H., & Mark, P. (1979). A Model for the Operation of a Thin-Film SnO x Conductance-Modulation Carbon Monoxide Sensor. Journal of The Electrochemical Society, 126(4), 627–633. https://doi.org/10.1149/1.2129098

Woodall, G. M., Hoover, M. D., Williams, R., Benedict, K., Harper, M., Soo, J.-C., … Thompson, B. (2017). Interpreting Mobile and Handheld Air Sensor Readings in Relation to Air Quality Standards and Health Effect Reference Values: Tackling the Challenges. Atmosphere, 8(10), 182. https://doi.org/10.3390/atmos8100182

Wu, S., Mickley, L. J., Leibensperger, E. M., Jacob, D. J., Rind, D., & Streets, D. G. (2008). Effects of 2000–2050 global change on ozone air quality in the United States. Journal of Geophysical Research: Atmospheres, 113(D6). https://doi.org/10.1029/2007JD008917

Yokelson, R. J., Burling, I. R., Gilman, J. B., Warneke, C., Stockwell, C. E., Gouw, J. de, … Weise, D. R. (2013). Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires. Atmospheric Chemistry and Physics, 13(1), 89–116. https://doi.org/10.5194/acp-13-89-2013

Yoo, K. S. (2011). Gas Sensors for Monitoring Air Pollution. https://doi.org/10.5772/19967

Zhang, L., Tian, F., Liu, S., Guo, J., Hu, B., Ye, Q., … Feng, J. (2013). Chaos based neural network optimization for concentration estimation of indoor air contaminants by an electronic nose. Sensors and Actuators A: Physical, 189, 161–167. https://doi.org/10.1016/j.sna.2012.10.023

Zhang, S., Lei, T., Li, D., Zhang, G., & Xie, C. (2014). UV light activation of TiO2 for sensing formaldehyde: How to be sensitive, recovering fast, and humidity less sensitive. Sensors and Actuators B: Chemical, 202(Supplement C), 964–970. https://doi.org/10.1016/j.snb.2014.06.063

Ziemann, P. J., & Atkinson, R. (2012). Kinetics, products, and mechanisms of secondary organic aerosol formation. Chemical Society Reviews, 41(19), 6582–6605. https://doi.org/10.1039/C2CS35122F

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., … Subramanian, R. (2017). Closing the gap on lower cost air quality monitoring: machine learning calibration models to improve low-cost sensor performance. Atmos. Meas. Tech. Discuss., 2017, 1–36. https://doi.org/10.5194/amt-2017-260

Zimmerman, Naomi, Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., … Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques; Katlenburg-Lindau, 11(1), 291–313. http://dx.doi.org/10.5194/amt-11-291-2018

APPENDIX A: SENSORS AND THEIR ASSOCIATED VARIABLE NAMES

Sensor values were named as indicated below in Table 5. Table of sensors that were used for testing and the corresponding variable name(s) associated with those sensors. Two sensor "pods" were joined together, so the name of the pod was appended to the end of the parameter name that is listed below. For example, the primary temperature sensor on pod E2 was named "temperatureYPODE2". The BME 180, SHT25, and ELT CO_2 sensors all communicated their output values to the control board via I^2C protocol. All other sensors were recorded as analog values using ADCs to record the induced voltage. An additional variable was also added ("telapsed") that indicates the time that the pod was turned on and is in the units of days. This variable is important as it indicates time that the sensors were at their operating conditions and would allow models to account for sensor drift caused by operating for an extended period of time.

Table 4. Table of sensors that were used for testing and the corresponding variable name(s) associated with those sensors.

Sensor Details	**Variable Name(s)**
Bosch BME 180 Barometric Pressure Sensor	bme_P
Sensirion Temperature and Humidity Sensor	temperature, humidity
ELT S300 NDIR CO2 Sensor	co2_NDIR
Alphasense NO-B4 Sensor, Working and Auxiliary Electrode	NO_B4_aux, NO_B4_main
Alphasense NO2-B1 Sensor, Working and Auxiliary Electrode	NO2_B1_aux, NO2_B1_main
Alphasense CO-B4 Sensor, Working and Auxiliary Electrode	CO_B4_aux, CO_B4_main
Alphasense H2S-BH Sensor, Working and Auxiliary Electrode	H2S_BH_aux, H2S_BH_main

Sensor Details	**Variable Name(s)**
Alphasense O3-B4 Sensor, Working and Auxiliary Electrode	O3_B4_aux, O3_B4_main
Figaro 2600 Installed on the Pod Board in the 210 and 280 mW Heater Circuits, Sensing Voltage	fig2600_210ob_s, fig2600_280ob_s
Figaro 2602 Installed on the Pod Board in the 280 mW Heater Circuit, Sensing Voltage	fig2602_280ob_s
Baseline Mocon PID Sensor	bl_mocon
MICS-5121wp Installed on the Pod Board, Sensing Voltage	MICS5121wp_ob_s
MICS-2611 Installed on the Pod Board, Sensing Voltage	e2v2611_ob_s
MICS-2611 Installed on the External Board, Sensing Voltage	mics2611_s
MICS-2710 Installed on the External Board, Sensing Voltage	mics2710_s
MICS-5525 Installed on the External Board, Sensing Voltage	mics5525_s
MICS-5121wp Installed on the External Board, Sensing Voltage	mics5121wp_s
Figaro 4161 Installed on the External Board, Sensing Voltage	fig4161_sens
Figaro 2600 Installed on the External Board, Sensing Voltage	fig2600_sens
Figaro 2611 Installed on the External Board, Sensing Voltage	fig2611_sens
Figaro 2602 Installed on the External Board, Sensing Voltage	fig2602_sens

APPENDIX B: FULL LIST OF TEST POINTS

Table 5. Full list of test points with the mean values of temperature, humidity, and concentration that were recorded during that test point. The "Total" column indicates the total concentration of gases in the chamber, not including dilution gas. Some testing was limited by gas availability, which is why the quantity of test points for each source and concentration are not consistent.

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
No Source	23	47	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	26	48	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	27	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	27	38	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	28	49	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	29	47	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	29	46	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	30	49	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	30	39	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	30	49	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	30	54	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	30	14	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	49	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	39	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	52	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	31	68	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	32	57	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	33	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
No Source	33	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	33	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	33	72	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	34	45	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	35	58	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	35	55	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	38	40	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	40	41	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	42	51	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	43	53	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
No Source	43	62	0.00	0.00	0.00	0	0.0	0.00	0.00	0.00
Gasoline Vapor	31	40	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	31	40	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	36	40	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	39	41	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	40	40	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	40	42	0.04	0.00	0.00	0	0.0	0.00	0.00	0.04
Gasoline Vapor	31	40	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	32	40	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	38	40	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	39	41	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	40	41	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	40	40	0.11	0.00	0.00	0	0.0	0.00	0.00	0.11
Gasoline Vapor	31	40	0.24	0.00	0.00	0	0.0	0.00	0.00	0.24
Gasoline Vapor	32	40	0.24	0.00	0.00	0	0.0	0.00	0.00	0.23
Gasoline Vapor	40	40	0.23	0.00	0.00	0	0.0	0.00	0.00	0.23

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
Gasoline Vapor	40	41	0.24	0.00	0.00	0	0.0	0.00	0.00	0.24
Gasoline Vapor	40	40	0.24	0.00	0.00	0	0.0	0.00	0.00	0.24
Gasoline Vapor	40	41	0.24	0.00	0.00	0	0.0	0.00	0.00	0.23
Gasoline Vapor	26	49	0.35	0.00	0.00	0	0.0	0.00	0.00	0.35
Gasoline Vapor	27	50	0.30	0.00	0.00	0	0.0	0.00	0.00	0.30
Gasoline Vapor	29	46	0.31	0.00	0.01	0	0.0	0.00	0.00	0.30
Gasoline Vapor	29	49	0.35	0.00	0.00	0	0.0	0.00	0.00	0.35
Gasoline Vapor	31	50	0.30	0.00	0.00	0	0.0	0.00	0.00	0.29
Gasoline Vapor	31	40	0.31	0.00	0.00	0	0.0	0.00	0.00	0.10
Gasoline Vapor	29	47	0.35	0.00	0.01	0	0.0	0.00	0.00	0.35
Gasoline Vapor	31	40	0.44	0.00	0.00	0	0.0	0.00	0.00	0.44
Gasoline Vapor	38	65	0.44	0.00	0.00	0	0.0	0.00	0.00	0.44
Gasoline Vapor	32	40	0.48	0.00	0.00	0	0.0	0.00	0.00	0.48
Gasoline Vapor	40	41	0.48	0.00	0.00	0	0.0	0.00	0.00	0.48
Gasoline Vapor	40	40	0.48	0.00	0.00	0	0.0	0.00	0.00	0.48
Gasoline Vapor	42	53	0.79	0.00	0.00	1	0.0	0.00	0.00	0.10
Heavy Exhaust	37	73	57.87	0.69	0.56	57	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	60.07	0.69	0.57	59	0.0	0.00	0.00	0.00
Heavy Exhaust	41	67	145.92	1.40	0.27	144	0.0	0.00	0.00	0.00
Heavy Exhaust	43	62	145.94	0.33	1.14	144	0.0	0.00	0.00	0.00
Heavy Exhaust	32	40	145.99	0.33	1.15	145	0.0	0.00	0.00	0.00
Heavy Exhaust	42	61	146.08	1.40	0.56	144	0.0	0.00	0.00	0.00
Heavy Exhaust	43	62	146.19	0.69	1.14	144	0.0	0.00	0.00	0.00
Heavy Exhaust	32	40	146.37	0.69	1.15	145	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	146.87	1.40	0.57	145	0.0	0.00	0.00	0.00
Heavy Exhaust	40	69	146.91	1.40	1.14	144	0.0	0.00	0.00	0.00

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
Heavy Exhaust	31	40	147.18	1.40	0.28	146	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	148.38	1.40	1.15	146	0.0	0.00	0.00	0.00
Heavy Exhaust	38	55	195.05	0.69	0.57	194	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	195.53	0.69	0.57	194	0.0	0.00	0.00	0.00
Heavy Exhaust	39	40	195.51	0.69	0.57	194	0.0	0.00	0.00	0.00
Heavy Exhaust	33	61	281.17	0.33	1.15	280	0.0	0.00	0.00	0.00
Heavy Exhaust	39	48	282.55	1.40	0.28	281	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	282.86	1.40	0.28	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	46	282.91	1.40	0.57	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	44	282.90	0.33	1.15	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	45	283.02	0.69	1.15	281	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	283.13	0.69	1.15	281	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	283.17	1.40	0.57	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	51	283.28	1.40	1.15	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	40	283.46	1.40	0.28	282	0.0	0.00	0.00	0.00
Heavy Exhaust	38	40	283.56	0.33	1.15	282	0.0	0.00	0.00	0.00
Heavy Exhaust	31	40	283.86	1.40	1.15	281	0.0	0.00	0.00	0.00
Heavy Exhaust	39	40	283.92	0.69	1.15	282	0.0	0.00	0.00	0.00
Heavy Exhaust	39	40	283.94	1.40	0.57	282	0.0	0.00	0.00	0.00
Heavy Exhaust	39	40	284.37	1.40	1.15	282	0.0	0.00	0.00	0.00
Low T Combustion	29	49	0.44	0.00	0.44	0	0.0	0.00	0.00	0.00
Low T Combustion	30	49	0.45	0.00	0.45	0	0.0	0.00	0.00	0.00
Low T Combustion	30	50	0.36	0.00	0.36	0	0.0	0.00	0.00	0.00
Low T Combustion	31	40	0.45	0.00	0.45	0	0.0	0.00	0.00	0.00
Low T Combustion	40	60	0.45	0.00	0.45	0	0.0	0.00	0.00	0.00
Low T Combustion	31	40	103.54	0.00	0.00	104	0.0	0.00	0.00	0.00

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
Low T Combustion	36	60	103.78	0.00	0.00	104	0.0	0.00	0.00	0.00
Low T Combustion	38	40	103.93	0.00	0.00	104	0.0	0.00	0.00	0.00
Low T Combustion	39	40	104.02	0.00	0.00	104	0.0	0.00	0.00	0.00
Low T Combustion	30	49	564.64	0.00	0.00	565	0.0	0.00	0.00	0.00
Low T Combustion	41	56	567.67	0.00	0.01	568	0.0	0.00	0.00	0.00
Low T Combustion	31	40	570.85	0.00	0.01	571	0.0	0.00	0.00	0.00
Natural Gas	30	40	1.49	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	30	39	1.49	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	31	40	1.50	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	34	40	1.49	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	34	45	1.49	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	40	58	1.51	0.00	0.00	0	1.5	0.00	0.00	0.00
Natural Gas	30	39	1.92	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	30	40	1.92	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	31	40	1.92	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	32	56	1.92	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	32	52	1.92	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	34	40	1.93	0.00	0.00	0	1.5	0.43	0.00	0.00
Natural Gas	42	55	1.92	0.00	0.00	0	1.5	0.42	0.00	0.00
Natural Gas	30	40	3.29	0.00	0.00	0	3.3	0.00	0.00	0.00
Natural Gas	30	39	3.28	0.00	0.00	0	3.3	0.00	0.00	0.00
Natural Gas	34	40	3.28	0.00	0.00	0	3.3	0.00	0.00	0.00
Natural Gas	34	46	3.27	0.00	0.00	0	3.3	0.00	0.00	0.00
Natural Gas	35	47	3.28	0.00	0.00	0	3.3	0.00	0.00	0.00
Natural Gas	30	40	4.18	0.00	0.00	0	3.3	0.90	0.01	0.00
Natural Gas	31	39	4.18	0.00	0.00	0	3.3	0.90	0.01	0.00

Source	T (C)	RH (%)	Total (ppm)	NO_2 (ppm)	CO (ppm)	CO_2 (ppm)	CH_4 (ppm)	Ethane (ppm)	Propane (ppm)	Gasoline (ppm)
Natural Gas	33	52	4.18	0.00	0.00	0	3.3	0.90	0.01	0.00
Natural Gas	33	50	4.18	0.00	0.00	0	3.3	0.90	0.01	0.00
Natural Gas	34	40	4.19	0.00	0.00	0	3.3	0.90	0.01	0.00
Natural Gas	30	40	13.93	0.00	0.00	0	13.9	0.01	0.00	0.00
Natural Gas	30	39	13.92	0.00	0.00	0	13.9	0.00	0.00	0.00
Natural Gas	31	40	13.92	0.00	0.00	0	13.9	0.01	0.00	0.00
Natural Gas	34	40	13.94	0.00	0.00	0	13.9	0.01	0.00	0.00
Natural Gas	34	47	13.93	0.00	0.00	0	13.9	0.01	0.00	0.00
Natural Gas	35	48	13.94	0.00	0.00	0	13.9	0.00	0.00	0.00
Natural Gas	38	62	13.91	0.00	0.00	0	13.9	0.00	0.00	0.00
Natural Gas	31	40	17.65	0.00	0.00	0	13.9	3.74	0.05	0.00
Natural Gas	43	52	17.64	0.00	0.00	0	13.9	3.72	0.05	0.00
Natural Gas	30	40	17.68	0.00	0.00	0	13.9	3.74	0.07	0.00
Natural Gas	30	39	17.71	0.00	0.00	0	13.9	3.75	0.07	0.00
Natural Gas	34	48	17.68	0.00	0.00	0	13.9	3.74	0.07	0.00
Natural Gas	34	40	17.69	0.00	0.00	0	13.9	3.74	0.07	0.00
Natural Gas	34	49	17.69	0.00	0.00	0	13.9	3.74	0.07	0.00

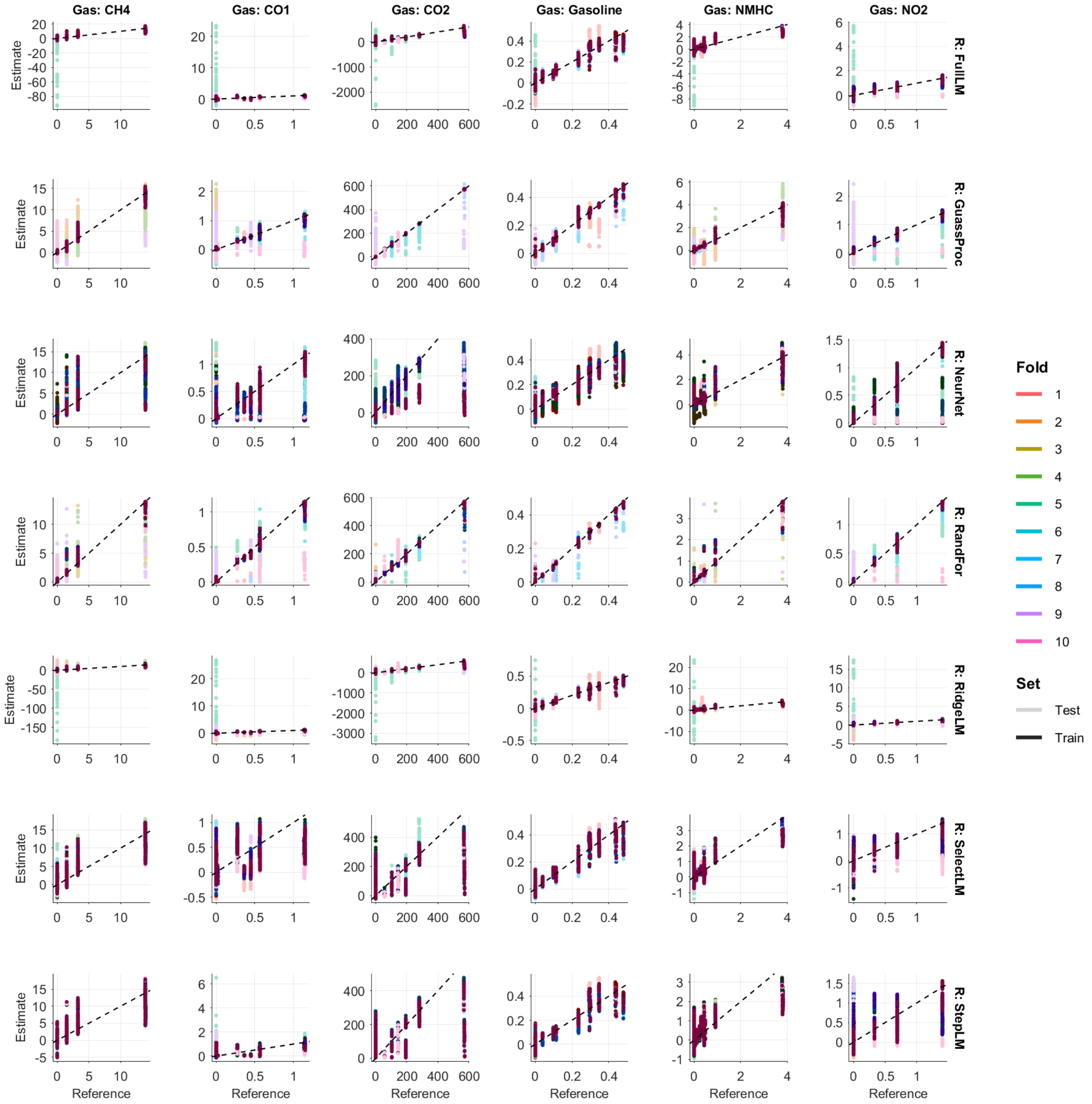


Figure 14. Estimated versus reference concentrations are plotted on the next page for each combination of gas and regression technique. Each column of plots contains estimates for a given gas, and each row contains estimates for a given regression technique. The color of each point indicates the cross validation fold that was used for training and testing the model, and the shade (light versus dark) indicate whether the values were estimated on training data or testing (validation) data.