Development and Characterization of a Novel, Low-Cost Method for Measurement of Volatile Organic Compounds

by

Amanda Gao

B.S. Chemical Engineering and English, California Institute of Technology, 2018

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

© 2024 Amanda Gao. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by:	Amanda Gao Department of Civil and Environmental Engineering April 25, 2024
Certified by:	Jesse H. Kroll Professor of Civil and Environmental Engineering, Thesis Supervisor
Accepted by:	Heidi M. Nepf Donald and Martha Harleman Professor of Civil and Environmental Engineering Chair, Graduate Program Committee

Development and Characterization of a Novel, Low-Cost Method for Measurement of Volatile Organic Compounds

by

Amanda Gao

Submitted to the Department of Civil and Environmental Engineering
on April 25, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Measurements of atmospheric pollutants are crucial for improving our understanding of atmospheric chemistry, managing air quality, and estimating exposure to compounds that have profound impacts on human health. Low cost sensors (LCS), due to order-of-magnitude reductions in power usage, maintenance needs, and purchase cost compared to research-grade reference instruments, have the potential to greatly expand the spatiotemporal resolution of these measurements. While there are several commercially-available LCS that can measure environmental volatile organic compounds (VOCs), an important class of hazardous pollutants, these sensors can only make non-specific "broadband" measurements and have, to date, been underutilized in research.

This thesis describes the development, characterization, optimization, and use of a novel low-cost instrument for measuring environmental VOCs. This instrument utilizes an array of low-cost VOC sensors representing three fundamentally different sensor types. It also takes advantage of user-controlled parameters that achieve greater degrees of differentiation between responses of sensors with the same measurement type. In the first part of this work, we describe the instrument itself, as well as a laboratory study that characterizes sensor responses to environmentally relevant VOCs. Though environmental applications pose unique challenges that can't be completely addressed in the laboratory, our results demonstrate that this instrument can give quantitative, chemically specific information about VOCs.

The second part of this work is based on measurements made as part of a collaborative indoor air quality campaign, where our low-cost VOC instrument and co-located reference monitors made measurements of realistic indoor VOC sources. Results from an LCS-derived matrix factorization analysis were compared to an independent factor analysis of reference VOC measurements, demonstrating that our uncalibrated low-cost data can provide quantitative and qualitative information about VOC sources and composition. Based on this comparison analysis, we describe a procedure for sensor selection that allows us to evaluate the relative importance of specific sensors or sensor types in providing information about VOC composition and sources, helping future similar LCS array applications to avoid measurement redundancies and minimize material cost.

Overall, the results from this thesis show that this LCS instrument can provide useful, quantitative information about VOC sources and composition at a fraction of the size and cost of a research-grade instrument–opening the possibility of widespread and spatially distributed measurements of VOCs in air quality and chemistry contexts, especially for indoor air.

Thesis supervisor: Jesse H. Kroll
Title: Professor of Civil and Environmental Engineering

Acknowledgments

I'm really lucky to have been advised by the best boss ever, Jesse Kroll. It's rare that an advisor is both highly empathetic and highly intelligent and uses both attributes to their full effect: every weekly meeting with Jesse left me feeling more energized about future research and my own progress, no matter how beat down I felt before entering his office or Zoom room. I would also like to thank my committee members, Professors Colette Heald and David Hsu, for their commitment to helping me improve my research and for their extremely helpful input during committee meetings. They are both very empathetic and clever people, and I really appreciate how even their off-the-cuff comments led my research to some unexpectedly intriguing and insightful directions.

Even though I worked on a "black sheep" low-cost sensor project, I somehow could not avoid the highly collaborative nature of atmospheric chemistry. During my time with the Kroll group, I've crossed paths with many graduate students and postdocs (Chris, Ben, Kevin, Abby, David, Mark, Josh, James, Qing, Tori, Amy, Matthew, Erik, Nadia, Lesly, Hannah, Lexy, Seamus, Jia, and Isabel) who have all, by some miracle, been very pleasant, helpful and whip-smart people who have been a joy to work with. In particular, I'd like to thank my former labmate David Hagan for all sorts of help with low-cost sensors throughout the years, as well as my labmates Matthew Goss and Erik Helstrom for putting aside their own research temporarily to help design and assemble the low-cost sensor arrays. Thanks to mini-UROP student Iselle Barrios, who helped to run and interpret preliminary lab characterization tests. I'd also like to thank former Kroll group postdocs Rachel O'Brien and Amy Hrdina, who ensured that our sensors stayed alive during CASA.

This work would not be possible without contributions from all the researchers at CASA, especially Professors Nina Vance and Delphine Farmer and NIST staff scientist Dustin Poppendieck. I'd also like to thank the Abbatt group at UToronto, especially Han Huynh and Jenna Ditto, for supplying us with reference data and helping to interpret our analysis results. A final thanks goes the Isaacman-VanWerz Group at Virginia Tech and the Goldstein group at U.C. Berkeley for all the effort put into helping us to make unsuccessful measurements.

Finally, I have to acknowledge everybody who supported me outside of a strictly academic context. I cherish all the sweet and sometimes crazy memories from being a graduate resident advisor at AXO and WILG, and I really appreciate the many excellent undergraduate students I met through the years. Also thanks to my friends both in the Boston area and far away: I always knew I could count on you all for support and good experiences. Thanks to Jules, the best workout/climbing buddy and supplier of tens of thousands of calories in hot Cheetos and gummy bears that were converted to the words you're reading right now. I'd also like to thank my parents and siblings for being so unconditionally supportive, even if they haven't quite yet figured out what exactly I do all day as a PhD student... honestly I'm not quite sure either, but I guess it doesn't matter now...

Contents

List of Figures

List of Tables

Commonly-Used Abbreviations

VOC	Volatile Organic Compound
PM	Particulate Matter
LCS	Low-Cost Sensor
PID	Photo-Ionization Detector
EC	Electrochemical (Sensor)
MOx	Metal Oxide
NMF	Non-negative Matrix Factorization
RH	Relative Humidity
T	Temperature

Chapter 1

Introduction

1.1 The Importance of Low-Cost Atmospheric Measurements

Measurements of atmospheric pollutants are crucial for improving our understanding of atmospheric chemistry, managing air quality, and estimating exposure to compounds that negatively impact human health. Exposure to air pollution has a profound societal impact, as it remains a leading risk for attributable deaths and shortened lifespans worldwide [1]. Even the most conservative estimates attribute around 4 million annual excess deaths to air pollution, with other methodologies attributing over 10 million annual excess deaths [2]. A handful of compounds have an outsized impact on human health, including particulate matter (PM) and ozone, which are associated with respiratory diseases [3], cardiovascular diseases [4], and adverse birth outcomes [5], [6]. Many of these important air contaminants are secondary, meaning they form through chemical reactions involving precursor compounds: for example, ground-level ozone and PM can form from the reaction of NOx and volatile organic compounds (VOCs) in the presence of sunlight [7]. Thus, obtaining the complete picture of air quality and composition requires measurements of both hazardous pollutants and their key precursors.

Measurements of harmful indoor pollutants and their precursors are particularly important, as humans generally spend the majority of their time indoors [8]; moreover, exposure to household air pollution contributes to multiple adverse health effects such as childhood asthma and low birth weight [9], [10]. In fact, exposure to indoor air pollution ranks among the top ten risk factors for disease worldwide [11]. Indoor pollutants can originate from building materials and static contents [12], and are also emitted from common household activities such as cooking or application of personal care products [13]. Indoor and outdoor air quality and chemistry are also inextricably linked: intrusion of outdoor pollutants is a major contributor to indoor air quality [14], while indoor activities such as cooking and wood burning are significant sources of outdoor air pollution [15]. Traditionally, real-time measurements of indoor and outdoor atmospheric pollutants are made using reference instruments with high precision and accuracy that can be prohibitively expensive in terms of material cost and operating requirements. The high cost of these instruments contributes to inequities in measuring air pollution: for example, air pollution disproportionately impacts low-and middle-income countries [1], yet these regions are also the most likely to have an air quality data gap [16]. Even regions with well-developed monitoring infrastructure struggle to systematically measure the smaller, sub-regional differences in exposure (e.g. variations across “micro-environments” such

as the home, office, or transit) that have significant effects on personal exposure and risk [17].

To fill these knowledge gaps, many researchers and regulatory bodies have turned to measurements of atmospheric pollutants made by low-cost sensors (LCS). The "low-cost" descriptor is relative, and generally applies to sensors that have a purchase cost at least one order of magnitude lower than that of a reference instrument measuring the same pollutant [18]. These sensors have seen major technological improvements in the last two decades that enable them to measure ambient levels of atmospheric pollutants in the parts-per-billion (ppb) range [19]. In addition to their low cost, LCS also have the added benefits of occupying very little physical space, drawing little power, and generally not requiring human intervention to operate. The high spatiotemporal resolution of LCS measurements makes them good candidates for expanding our knowledge of air quality and chemistry via novel applications such as distributed sensor networks, [20] personal exposure measurements [21], and sensor arrays that make co-located, multi-pollutant measurements [22].

1.2 Low-Cost VOC Measurements

LCS have been extremely helpful in characterizing the behavior of key inorganic pollutants, such as PM2.5 [23], O_3 [24], and SO_2 [25], but low-cost measurements of atmospheric volatile organic compounds (VOCs) remain relatively rare [26]. VOCs are an important class of atmospheric pollutants which are emitted from numerous natural sources and human activities [27]. Exposure to VOCs can be directly harmful to human health [28], [29] and emitted VOCs can also form hazardous secondary products, including peroxides [30], ozone [31], and fine PM in the form of secondary organic aerosol (SOA) [7]. Online, real-time measurement of VOCs was made possible by the relatively recent development of novel measurement methods, such as proton transfer reaction mass spectrometry (PTR-MS) [32] and chemical ionization mass spectrometry (CIMS) [33]. Although these technologies demonstrate a significant improvement over extant off-line technologies for measuring VOCs, such as the use of sorbent tubes to trap VOCs for manual GC-MS analysis [34], instruments employing these state-of-the-art measurement techniques are even more prohibitively enormous, energy-consumptive, and expensive than the typical reference monitor. Hence, a feasible LCS alternative would open the possibility of widespread and spatially distributed measurements of VOCs in air quality and chemistry contexts.

While there are several different types of comercially-available LCS that can measure VOCs at ambient, parts-per-billion (ppb) concentrations, they are limited by their non-specific (“broadband”) nature: any individual sensor can only output a single scalar value that reflects a combination of different sensor sensitivities toward a wide and poorly-defined range of VOCs [26]. A potential solution to this problem is to simply use more than one individual VOC sensor: in theory, meaningful differences between an array of partially selective low-cost VOC sensor responses can be leveraged, via a pattern recognition algorithm, to gain useful information about the measured compound or mixture. This approach has been the linchpin of “electronic nose” studies that mostly aim to detect and classify VOC mixtures in odor detection or process control applications [35]. However, environmental VOCs pose a particular challenge for these applications: most “electronic noses” are designed to measure VOCs at concentrations (generally tens or hundreds of parts-per-million, or ppm) that are many orders of magnitude higher than the ppb levels found in the atmosphere [35], [36]. The challenge of these measurements is further exacerbated by the complexity of atmospheric VOC sources, compositions, and variations [37]–[39] as well as sensor cross-sensitivities to non-VOC

gases and environmental parameters such as relative humidity and temperature [26].

A handful of past sensor array applications have attempted to obtain quantitative measurements of sub-ppm VOC pollution sources, sidestepping the challenges posed by environmental VOC complexity by directing their focus toward quantifying a single significant VOC. For example, methane has been an important target for several sensor arrays [40]–[42] that each combined multiple metal oxide sensors and a machine learning regression algorithm to estimate variations in environmental methane concentrations. Benzene is another important VOC that LCS have high sensitivity to, and De Vito et al. showed that ambient concentrations of benzene measured near a major Italian roadway could be accurately estimated by a neural network model trained on measurements from five different metal oxide sensors [43]. Similarly, Collier-Oxandale et al. used measurements from two different metal oxide sensors to develop regression models for atmospheric benzene, methane, and total VOC concentrations measured near Denver, Colorado [44]. In sum, these studies have shown that an array of metal oxide sensors can be successfully used as inputs for regression models that give quantitative estimates of certain VOC concentrations.

While these past studies have shown the potential of LCS sensor arrays to generate quantitative, chemically specific VOC information, they are somewhat limited in scope. Previous studies utilized only one measurement technology (metal oxide sensing) to make environmental measurements. Multiple other low-cost measurement technologies with sub-ppb VOC sensitivities are commercially available [26], and their potential to contribute to sensor array measurements of environmental VOCs is completely unexplored. In addition, these studies have focused on developing calibration models for a single VOC of interest, but we believe that LCS array measurements could potentially be leveraged for more insightful data on VOC sources and composition. A simple example comes from Collier-Oxandale et al., who noticed that differences in signal between the two low-cost VOC sensors in their array gave some information on chemical sources—namely, that the ratio of the two metal oxide responses, due to the sensors' varying sensitivities to toluene and benzene, could potentially be used to separate measurements of traffic from measurements of oil-and-gas emissions [44]. It seems feasible, then, that data from a larger sensor array incorporating multiple sensing technologies could provide even more information on the underlying sources of VOCs and other pollutants.

Some past studies have recognized that low-cost VOC sensors could be useful in making spatiotemporally distributed measurements of indoor air, especially given that the higher indoor concentrations of VOCs [45] partially compensate for the sensors' relatively higher limits of detection. Unfortunately, these studies usually struggled to transfer laboratory results to the field: Zhang et al. [46] and Wolfrum et al. [47] developed metal oxide sensor arrays for indoor VOC monitoring and showed that these arrays could classify common indoor VOCs at high concentrations within the laboratory, but did not extend their analyses to real indoor air. Arnold et al. [48] developed a metal oxide microarray that was able to classify some common VOCs at high concentrations in the laboratory and showed that the array's responses to two real-world pollution sources (smoldering cables and an occupied meeting room) could be clearly distinguished from the its responses to isopropanol-soaked tissues. However, the high concentrations of the tested compounds and sources suggest that this particular approach is only suited for identifying extreme VOC emission events rather than typical indoor pollution sources. In summary, while there have been several attempts to develop a low-cost indoor VOC monitor using laboratory calibrations, to our knowledge there have been no attempts to measure realistic and common indoor VOC pollution sources in the field with a low-cost sensor array.

1.3 Objectives for this Thesis

In this thesis, we describe the development and application of a novel instrument, consisting of an LCS array, that measures environmental VOCs. Unlike previous sensor arrays developed for similar applications, our array of twelve sensors utilizes three different low-cost sensing technologies: metal-oxide sensing, photo-ionization detection, and electrochemical (amperometric) sensing. We also achieve additional differences in VOC sensitivity by varying operational parameters between otherwise identical sensors, a technique that has generally been underutilized in sensor array applications and has not been explored as a method for measuring environmental VOCs. Rather than attempting to use sensor outputs to train regression models for one or several individual VOCs like previous sensor array applications, we instead investigate the potential of this array to yield physically interpretable, chemically specific, and quantitative information about multiple groups of VOCs representing emissions from different pollution sources.

In Chapter 2, we describe the development and design of the low-cost VOC instrument, including descriptions of the custom hardware and firmware that we implemented. We report the results of a laboratory characterization study that subjected the instrument to a realistic concentration range of 10 atmospherically relevant VOCs. We investigated and quantified the effects of relative humidity and VOC mixtures on sensor responses. These laboratory results help us to understand the abilities and limitations of the array and to use this information to identify suitable applications for the low-cost instrument.

In Chapter 3, we describe measurements of indoor VOC sources made using our low-cost VOC instrument. Our low-cost VOC instrument and co-located reference monitors made measurements of various chemical perturbations representing realistic sources of VOCs, such as cooking, pesticide application, and wildfire smoke intrusion. Multi-pollutant low-cost measurements, including low-cost VOC measurements, were used to perform a factor analysis that identified periods of time when the air was influenced by different sources of indoor pollution. Results from the LCS-derived factorization were then compared to an independent source apportionment, performed using research-grade reference VOC measurements, that we show is highly representative of known VOC sources. Through this comparison analysis, we show that our low-cost data can provide quantitative and qualitative information about VOC sources and composition.

Finally, in Chapter 4, we describe a procedure for sensor selection to identify any sub-arrays that can provide the same amount of VOC information as the full array. To do this, we evaluate many possible sensor sub-arrays using a “brute force” search process and a simple scoring scheme based on the matrix factorization analysis developed in Chapter 3. These results allow us to evaluate the relative importance of specific sensors or sensor types in providing information about VOC composition and sources. We also show that the size of the LCS array can be significantly reduced, helping future similar LCS array applications to avoid measurement redundancies and minimize material cost.

In summary, this thesis carries out an in-depth investigation of the suitability of making measurements of environmental VOCs with low-cost sensors, using both laboratory and in-field data to assess sensor abilities and limitations. The results from this work suggest that low-cost sensors can’t provide nearly the same sensitivity or chemical detail expected from state-of-the-art, research-grade instruments. However, despite these limitations, low-cost VOC measurements can be used to better constrain human exposure to air pollution and to better understand the sources and trans-

formations of emitted VOCs. We also show that even a limited number of VOC sensors can be carefully chosen to maximize the information gained about VOC sources and composition. The technologies and approaches described in this work will help the future development of sensor array applications that provide spatially distributed, real-time measurements of VOCs and contribute to our fundamental understanding of chemical composition and human exposure across scales.

Chapter 2

Development and Characterization of a Low-Cost VOC Instrument

2.1 Introduction

In Chapter 1, we discussed the importance of VOCs, a class of atmospheric pollutants emitted from numerous natural sources and human activities [27] that can be directly harmful to human health [28] and contribute to the formation of hazardous secondary products [7]. A feasible low-cost alternative to the costly state-of-the-art measurement techniques would open the possibility of widespread and spatially-distributed measurements of VOCs in air quality and chemistry contexts. Several different types of LCS that can measure VOCs are commercially available, but it is challenging to make quantitative measurements with these sensors due to the inherent complexity of atmospheric VOC sources, compositions, and variations [37][38], as well as sensor cross-sensitivities to non-VOC gases and environmental parameters such as relative humidity and temperature[26]. Perhaps the most significant challenge of all is the non-specific or "broadband" nature of an individual low-cost VOC sensor, which can only output a single scalar value that reflects a combination of varying sensitivities toward a wide and poorly-defined range of VOCs [26].

This “broadband” nature of individual VOC sensors has been a long-standing problem in low-cost sensor research. So-called “electronic nose” instrument designs, originating in the late 1980s, attempt to overcome this inherent limitation by utilizing an “array of electronic chemical sensors with partial specificity and an appropriate pattern-recognition system” for recognizing simple or complex mixtures of chemical species [35]. These instruments were generally developed for odor detection and process control applications where concentrations of VOCs are generally many orders of magnitude higher than in the ambient environment [35]. Decades ago, researchers recognized that these “electronic noses” could be used to detect atmospheric pollutants, but such applications remained relatively uncommon due to the challenges of low ambient concentration and sensor cross-sensitivities with relative humidity and temperature [26].

Despite these challenges, there have been several past attempts to utilize sensor arrays to measure VOCs in the atmospherically relevant concentration range (ppb). Most of these “electronic unique nose” studies focused on a single pollutant or a limited subset of compounds. For example, De Vito et al. used an array of five bespoke metal oxide sensors, as well as a neural network model trained on sensor data gathered in the field, to accurately predict ambient benzene

concentrations in the range of 0.86 to 86 ppb measured near a major Italian roadway [43]. Wolfrum et al. recognized that a similar metal oxide array could have applications in indoor VOC measurements [47], where concentrations of certain compounds can be orders of magnitude higher than outdoors [45]. They attempted to leverage minute differences in response time and absolute sensitivity caused by manufacturing inconsistencies between 13 identical metal oxide sensors, and used the array of responses to obtain partial least squares calibrations for toluene, acetone, and isopropanol between 10-300 ppb [47]. More recently, Collier-Oxandale et al. utilized two different metal oxide sensors to obtain reasonable multi-linear models for predicting methane, benzene, and summed VOC concentrations at an observatory near Denver; they also noticed that the ratio of the two metal oxide responses, which vary due to varying sensitivities to toluene and benzene, could give some information on the VOC sources present i.e. distinguishing between traffic or oil-and-gas emissions [44].

In summary, there has been substantial research on low-cost sensor arrays, but only a small fraction of this work has focused on environmental monitoring. The studies that do focus on environmental applications have predominately been concerned with obtaining machine-learning calibration models for a single important environmental VOC or a small subset of them, generally at higher VOC concentrations where interfering effects are smaller. Moreover, these arrays have predominantly only relied on a single sensing technology: metal oxide sensing, a holdover from the very first "electronic nose" developed in 1982 [49]. To our knowledge, there has been no attempt to leverage multiple different sensing technologies to make sensor array measurements of atmospheric VOCs, nor has there been any investigation of adjusting operational parameters between otherwise identical sensors to obtain more varied sensitivities to these compounds.

Here, we describe the development of a novel, low-cost instrument for real-time measurement of ambient VOCs. This instrument utilizes a low-cost sensor array with 12 distinct low-cost sensors representing three different measurement technologies and takes advantage of user-controlled parameters that achieve greater degrees of differentiation between responses of sensors with the same measurement type. We also show laboratory characterization results for 10 key atmospheric VOCs in the 5 to 100 ppb range, as well as data from a wide range of relative humidities and a basic binary mixture. Finally, we will discuss these results in the context of practical usage of this array for environmental monitoring and evaluate the potential for this sensor array to provide useful, quantitative information about VOCs in realistic ambient conditions.

2.2 Methods

2.2.1 Low-Cost Sensing Principles

Our sensor array utilizes three different sensing technologies: metal oxide sensors (MOx), which measure target gas molecules that adsorb on a metal oxide surface; photo-ionization detectors (PID), which ionize gas molecules with a small vacuum ultraviolet lamp, and amperometric electrochemical (EC) sensors, which detect gases via an oxidation or reduction reaction in an electrolyte. Each of these measurement techniques has at least one adjustable parameter that can be leveraged to obtain different VOC sensitivities between otherwise identical sensors.

Metal oxide sensors have long been a popular choice for sensor array applications because of their particularly low material cost and relatively high sensitivity to VOCs [36]. MOx sensors

measure VOCs using adsorption: electrons are trapped by adsorbed molecules, and the resultant band-bending by these charged molecules changes the measured conductivity [50]. Due to the kinetics of these surface reactions and a balance on the availability of surface sites, MOx sensors exhibit a power-law dependency with the target gas concentration: the log of MOx response is linear with the log of analyte gas [50]. MOx sensors can vary in the materials or morphologies used for the semi-conducting sensing layer, which can greatly affect sensing properties: past studies on MOx sensor arrays, such as those by De Vito [43] or Collier-Oxandale [44], have relied on the use of sensors with manufactured differences (e.g. distinct semiconductor properties) to introduce distinctions in sensor sensitivity that can then be exploited computationally (e.g. through use of pattern recognition techniques). When using an array of identical sensors rather than fundamentally different ones, it is theoretically possible for a user to introduce significant differences in sensitivity by varying the operation temperature, or supply voltage, of each sensor, as this affects the relationship between sensor conductance and analyte gas partial pressure [50]. This technique has seen some success in electronic nose applications, such as when Liu et al. employed temperature modulation of MOx sensors to help detect excessive methanol emitted from liquors [51], but has never been applied in the context of environmental VOC measurement.

Photo-ionization detection, pioneered by James Lovelock to measure trace vapors in the atmosphere [52], relies on ionization of target molecules by a lamp to induce a measurable change in electric potential that is proportional to the concentration of target gas. In theory, an array of low-cost PID sensors, each containing miniature lamps of different VUV wavelengths, would be able to discriminate VOCs based on differences in ionization energy amongst the target species. Despite the potential usefulness of having such a diverse array of PIDs, low-cost PID specifications are limited by the state of the technology. Only certain wavelengths are commercially available, with 9.6 eV, 10.0 eV, and 10.6 eV lamps being common options. Certain detectors may also include technologies that increase PID sensitivity by reducing interference with water vapor and other interfering gases [53].

Electrochemical sensors, also known as amperometric sensors, rely on reduction-oxidation (redox) reaction between a target gas and an aqueous acid electrolyte. Due to their high sensitivity and selectivity, they have been widely used in low-cost air quality monitoring of major inorganic pollutants such as ozone and carbon monoxide [18]. On the manufacturing side, different VOC EC sensors may contain distinct chemical filters or catalysts that control selectivity and sensitivity to the measured compounds [24]. The sensitivities of EC sensors to various VOCs can be tweaked via the user's application of a bias voltage, or a potential difference between working and reference electrodes [24]. EC sensors that measure VOCs non-specifically have commonly been marketed for personal protection applications, but we were unable to find examples of VOC EC sensors with ppb-sensitivity being used in atmospheric or air quality contexts despite their commercial availability.

2.2.2 Instrument Design

Measurements were made using a custom-built sensor array for the continuous, real-time monitoring of ambient VOCs and environmental variables (temperature, relative humidity). A generalized schematic of the instrument is shown in Figure 2.1. Our array features 12 low-cost VOC sensors: three EC sensors (one Alphasense ETO-B1 sensor and two Alphasense VOC-B1 sensors with varying bias voltages), three PIDs (one IonScience MiniPID 2 10.0 eV, one IonScience MiniPID 2 10.6 eV, and one IonScience MiniPID HS 10.6 eV), and six metal oxide sensors (three Figaro

TGS2602 sensors and three Figaro TGS2600 sensors, each with different supply voltages). Table 2.1 summarizes the sensors used in this design and any user-controlled parameters that may have been applied. At the time of manufacture, the material cost of all 12 sensors was ~$2000 USD, with the vast majority of this cost made up by PID sensors (with an average cost of ~$ 540 USD). This total cost may be higher than a typical LCS application, but is still orders of magnitudes lower than the cost of a PTR-MS or CIMS instrument.

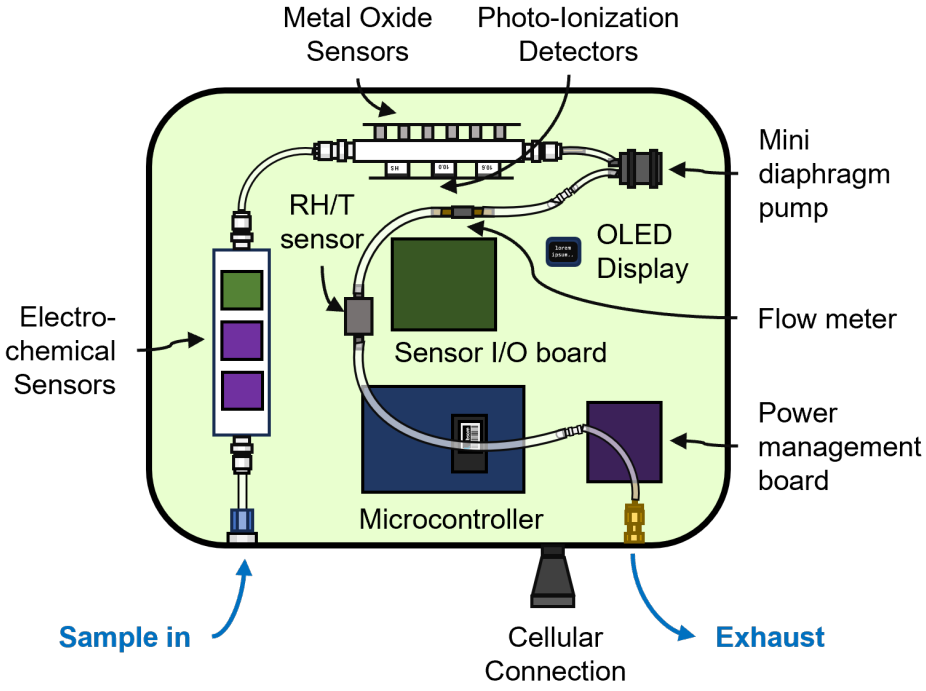

Figure 2.1 Schematic of the low-cost VOC sensor array. Sample is pulled in by a miniature diaphragm pump through custom flow cells that house 12 different VOC sensors. Several custom PCBs manage power and sensor inputs/outputs.

This is an active flow design, where air is drawn into the instrument by a miniature diaphragm pump (Xavitech v200) at a user-controlled rate that is generally set at 300 mL/min, but can be varied from 0-400 mL/min. Sample air travels through PFA tubing (1/4" outer diameter, 0.190" inner diameter), into custom-made Teflon flow cells (EC flow cell has dimensions 15.0 x 3.81 x 1.90 cm, MOx/PID flow cell has dimensions of 13.5 x 3.81 x 1.90 cm), with flow perpendicular to the sensor surfaces. After the sample air is expelled from the pump, it passes through a custom 3D-printed enclosure containing a relative humidity and temperature sensor (Sensiron SHT25) before being exhausted from the instrument. The design of this instrument maintains airtightness via O-rings that are flush against the sensor surfaces and mounting bolts that secure breakout circuit boards to the flow cells. As a result, sensors are not permanently secured to either the flow cells or their respective breakout circuit boards, allowing for easy replacement of any single sensor. The fully assembled instrument is housed inside a container with dimensions 42.2 x 37.1 x 21.0 cm (PolyCase ZH-161407) that is much larger than necessary to aid in troubleshooting this prototypical instrument.

The entire device is powered by mains electricity (12 V AC/DC converter) and controlled using

Table 2.1 Summary of VOC sensors used in sensor array instrument.

Sensor Model	Number of Sensors	Manufacturer	Sensing Technology	User-Applied Parameters
TGS 2600	3	Figaro Engineering, Inc.	MOx	Supply voltage (1 each at 4.75 V, 5.0 V, and 5.25 V)
TGS 2602	3	Figaro Engineering, Inc.	MOx	Supply voltage (1 each at 4.75 V, 5.0 V, and 5.25 V)
MiniPID 2 (10.0 eV)	1	ION Science Ltd.	PID	
MiniPID 2 (10.6 eV)	1	ION Science Ltd.	PID	
MiniPID 2 (10.6 eV)	1	ION Science Ltd.	PID	
ETO-B1 (10.6 eV)	1	Alphasense Ltd.	EC	
VOC-B4 (10.6 eV)	1	Alphasense Ltd.	EC	Bias Voltage (1 each at 0 mV, +300 mV)

an LTE-enabled microcontroller (Particle B Series SoM), which is utilized in conjunction with its manufacturer's evaluation circuit board (Particle M2EVAL). Several custom circuit boards manage sensor input and outputs, as well as associated analog-to-digital or digital-to-analog conversion; there is also a power management circuit board that supplies lines at 3.3 V, 5.0 V, and two variable values (intended for varying metal oxide supply voltages) that can be adjusted from 0.64V to 5.25V via user input to a synchronous buck regulator (MIC24045). Total power draw of the instrument is highest on startup, where the microcontroller alone requires ∼3 W. However, during regular operation most components have negligible power draw, but there are relatively large requirements from the pump (∼0.3 W), photo-ionization sensors (∼0.3 W), microcontroller (∼1 W), and metal oxide sensors (∼1.5 W) that result in < 5W of total power draw.

Data from all sensors is oversampled at ∼100 Hz before being averaged down to 1 Hz. This is substantially faster than some of the sensor response times, but oversampling helps to remove artifacts caused by electrical noise. The 1 Hz data is then logged to a local micro-SD card, and the 1-minute averages for the 1Hz data are computed and transmitted via 3G LTE to the cloud, where they are automatically processed and stored in Google Drive. The files for any single day's worth of data can then be downloaded or visualized in real time using a custom web app.

2.2.3 Laboratory Characterization

To achieve low and reliable concentrations of VOCs, we relied on headspace sampling, a technique where a volatile compound is placed in a sealed vial leaving sufficient room over the liquid, from which volatilized gas is sampled after phase equilibration. The concentration of the VOC in the gas phase at equilibrium can be calculated using the compound's temperature-dependent vapor pressure. For each experiment, we used a gas-tight syringe to obtain a headspace sample at 25 °C,

then placed the syringe into a computer-controlled syringe pump (Harvard Apparatus PHD Ultra), with the syringe needle inserted into a heated inlet maintained at 50 °C; the purpose of heating the inlet was to prevent condensation onto the tubing walls.

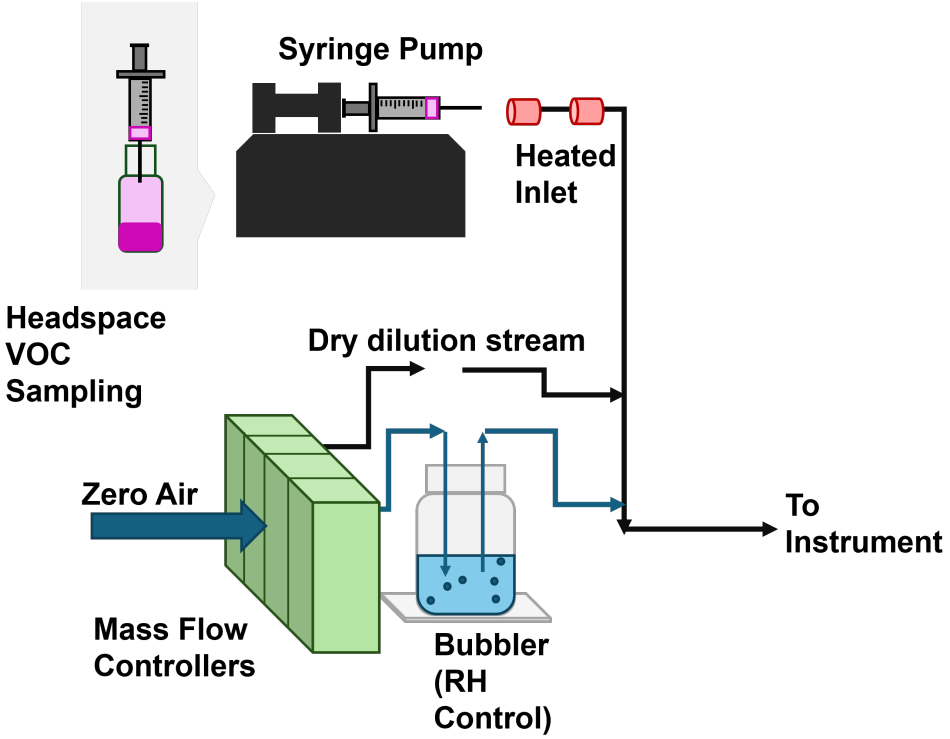

Figure 2.2 Schematic of the experimental apparatus for characterizing sensor responses to ppb-levels of VOCs. The VOC calibration gas system consists of gas-phase VOC obtained via headspace sampling that is then loaded into a gas-tight syringe and injected using a computer-controlled syringe pump. This is then diluted by a stream of zero air that may be humidified by a bubbler. The total flow is 10 lpm, but sample air is vented before reaching the instruments, preventing over-pressurization.

A dilution stream was also supplied to the inlet at 10 lpm, and MKS mass flow controllers were used to adjust the humidity of this dilution stream by varying the ratio of dry air line and a line humidified by a bubbler. The syringe pump was then operated with a preset, non-monotonic sequence of calibration levels, with each level being held for 25 minutes. To avoid any unwanted pressurization effects, the calibration mixture was vented before reaching the sensor array. Finally, after the calibration sequence was completed, the syringe was flushed with zero air several times before the next injection. A schematic diagram of the experimental setup for characterizing sensor array responses to ppb levels of VOCs is shown in Figure 2.2. An example calibration sequence, with sample sensor responses, is shown in Figure 2.3. The compound concentration is shown in the top panel with a dotted black line, while sensor responses are shown in the lower panels, along with the calibration curve and best fit that is derived from averaging steady-state sensor responses.

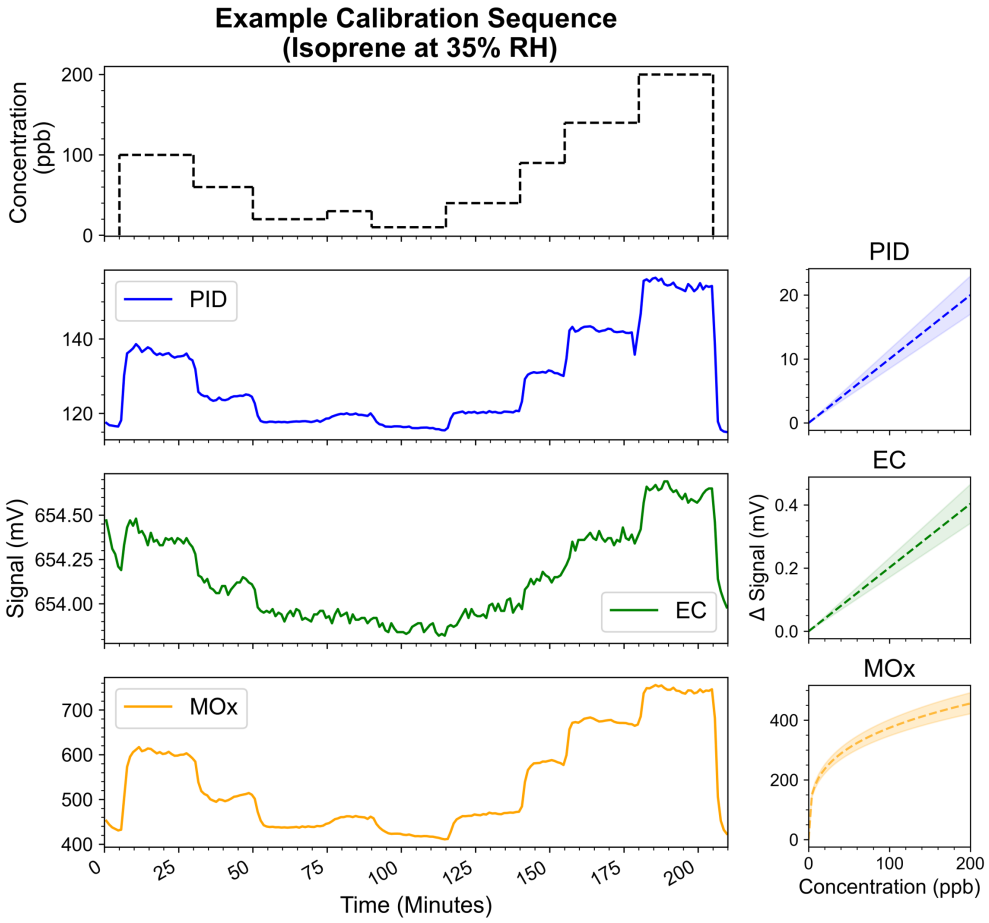

Figure 2.3 Example calibration sequence for a sample VOC (isoprene at 35% RH), with compound concentration shown in the top panel by the dotted black line. Sample responses from one PID, EC, and MOx sensor from the array are shown in the lower panels, while the right panels show calibration curves and best fits that can be derived from sensor responses.

2.2.4 Data Analysis

From each calibration sequence, we can average the steady-state values of sensor responses and derive a calibration curve, with signal as a function of concentration. We ignored the transient nature of these sensor responses: each concentration level was held for at least 20 minutes, and only the last 5 minutes of sensor responses were included in our analysis. Example calibration curves are shown in the right-hand panels of Figure 2.3. We found PID and EC curves to be linear, which is consistent with the results of prior studies [24], [54]. For these sensors, calibration points can be fit using a least-squares regression, where the slope represents a sensitivity in units of voltage per concentration unit. On the other hand, we noticed that MOx calibration curves were distinctly non-linear. We found that a power-law relationship describes these signals well, which is consistent with MOx physical sensing principles [50]. The observed responses can be expressed as $V = A[VOC]^{\beta} + C$, where [VOC] denotes the VOC concentration, A is an mV ppb$^{-\beta}$ sensitivity, β is a dimensionless power law parameter, and C is the signal baseline. Averaged MOx calibration points were fit to this expression using a non-linear least-squares regression. As a note, many

metal oxide sensor studies report signal in terms of a resistance ratio R/R_0, where R is the sensor resistance to a target gas and R_0 is the baseline resistance, but we have chosen to report output voltage to stay consistent with the other sensors in the array.[1]

Due to the challenge of characterizing baseline drifts, we characterize sensor responses in terms of net change in signal. In an ideal case, parameterization of sensor baseline variance could give information about the humidity and composition of background air [55], [56], but this is complicated by the additional dependencies of sensor baselines on electronic noise and sensor age [57]. We choose to prioritize characterization of sensor sensitivities rather than baselines as we expect that most practical applications of these sensors will involve baseline removal before data analysis. In our dataset, most baseline values were removed by a simple background subtraction. Some experiments showed mild drift between beginning and end values, usually caused by small fluctuations in relative humidity over the course of the experiment. In these situations, the baseline was identified and removed using the BaselineRemoval Python library (v1.0.5). The results from two different modified second-degree polynomial fits (ModPoly[58] and IModPoly[59]) were calculated, and the best of these methods was identified by minimizing calibration curve fit error.

2.3 Results and Discussion

2.3.1 Sensor Array Responses to Single VOCs

Sensor responses to 10 VOCs, broadly representative of those found in the atmosphere, were obtained in the range of 5 to 100 ppb at a constant relative humidity of 30% and a temperature of 22 °C. The sensor responses to these compounds (1-hexene, 1-octene, 2-pentanone, 2-heptanone, acetone, α-pinene, chlorobenzene, isoprene, o-xylene, and toluene) are summarized in Figure 2.4. Dotted lines represent the linear least squares regression of the measured values, denoted by triangles. In the range of concentrations tested, photo-ionization detectors (PIDs) and electrochemical sensors (EC) consistently exhibited a linear signal-to-concentration response. The high-sensitivity 10.6eV PID (ION Science MiniPID2) was able to detect all 10 compounds, with sensitivities ranging from $5.0\times10^{-2} \pm 1.5\times10^{-3}$ mV/ppb (o-xylene) to $1.76 \pm 4.6\times10^{-2}$ mV/ppb (acetone)[2]. Sensitivities for the other two PIDs (ION science MiniPID 10.0 and 10.6 eV sensors) were significantly lower, with a range of $6.0\times10^{-4} \pm 3.0\times10^{-4}$ to $3.3\times10^{-2} \pm 7.2\times10^{-3}$ mV/ppb for the 10.0eV PID, and $1.7\times10^{-3} \pm 3.0\times10^{-4}$ to $8.1\times10^{-2} \pm 2.0\times10^{-3}$ mV/ppb for the 10.6 eV PID.

Our array uses two different types of EC sensors, which were not able to detect the full suite of compounds and overall exhibited less sensitivity than the PID sensors. EC type 1 (Alphasense VOC-B1 EC sensor) detected six compounds, with sensitivities ranging from $6.6\times10^{-4} \pm 3.6\times10^{-4}$ mV/ppb (toluene) to $1.0\times10^{-2} \pm 1.2\times10^{-3}$ mV/ppb (isoprene). Application of a positive bias voltage generally increased sensor sensitivity to detected compounds: for example, the sensitivity of the biased sensor to isoprene was determined to be $2.7\times10^{-2} \pm 8.0\times10^{-3}$ mV/ppb, nearly triple the value of the unbiased sensor. However, biased sensor also had significant baseline drift, potentially

[1] The relationship between output voltage and the resistance ratio can be generally described with $R/R_0 \propto V_c/V_{out} - 1$, where V_{out} is the output voltage and V_c is the measurement circuit voltage, or 5 V for all our sensors. The conductivity G = 1/R is also sometimes used and should be roughly proportional to V_{out} based on the relationship outlined above.

[2] All confidence intervals reported in this chapter are $\pm 1\sigma$.

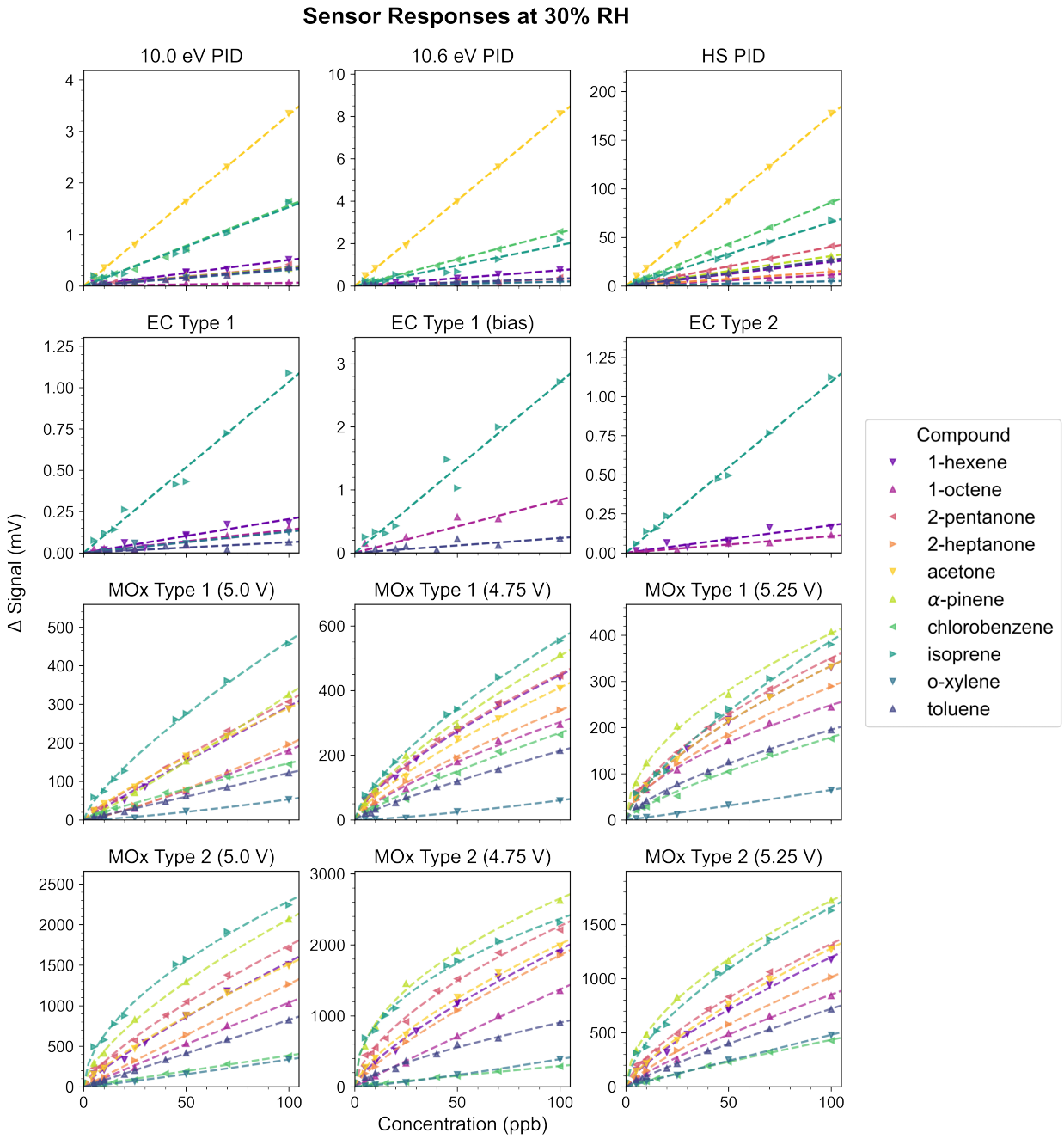

Figure 2.4 Summary of sensor responses (in mV over baseline) to various VOC concentrations between 5-100 ppb at 30% RH. Dotted lines represent the linear least squares regression of the measured values, denoted by triangles.

masking the sensor response to several compounds that the unbiased sensor was able to detect reliably (1-hexene and o-xylene). The second EC type in our array (Alphasense ETO-B1), measured four compounds with the sensitivities ranging from $1.1 \times 10^{-3} \pm 2.2 \times 10^{-4}$ mV/ppb (1-octene) to $1.1 \times 10^{-2} \pm 6.4 \times 10^{-4}$ mV/ppb (isoprene).

The metal oxide (MOx) sensors in our array detected all ten compounds with a power-law re-

sponse of voltage to concentration. Our array utilizes two different n-type metal oxide sensors (Figaro TGS2600 and 2602 models), which are both SnO_2 sensors but differ in the catalyst used in their sensing materials. These two models have very different sensitivity ranges to the VOCs tested: MOx type 1 (TGS2600) responses had a maximum observed sensitivity of 22 ± 5.1 mV $ppb^{-0.70\pm0.040}$ (isoprene), while MOx type 2 (TGS2602) saw a maximum of 140±29 mV $ppb^{-0.67\pm0.040}$(α-pinene). Further differences in sensitivity were achieved by varying the circuit voltage applied to individual sensors, which changes the operating temperature of each sensor. We observe that both the sensitivity A and the power law parameter β vary significantly between sensors, but we were unable to generalize the relationship between these parameters and applied voltage across all compounds. This finding is consistent with the results of Wang et al., who observed that the dependence of MOx VOC sensitivities on operating temperature is non-monotonic and compound-specific [55].

2.3.2 Sensor Array Responses as a Function of Relative Humidity

Environmental parameters are known to impact low-cost VOC measurements. Characterizing sensor responses to changing environmental temperature is beyond the scope of this study, but this effect could be important to fully understand responses of MOx sensors, which have a strong baseline signal dependence on ambient temperature [55], [60], [61]. EC baseline signals are also theoretically affected by temperature, as increasing temperature affects electrolyte density; however, this effect was found to be relatively minor for EC sensors utilizing H_2SO_4, such as the ones in our array [62]. PID sensitivities and compound ionization efficiencies do depend on sensor operating temperature, but operation under realistic ambient temperatures should not have a perceptible impact on sensor responses [63]. These prior studies suggest that sensor response dependence on environmental temperature should be relatively weak compared to the dependence on environmental RH which we will discuss here at length.

To explore the effects of RH on sensor responses, we obtained calibration curves for α-pinene and isoprene at many different relative humidity (RH) values. Figure 2.5, which shows PID calibration curves for isoprene at several different RH values, demonstrates the drastic effect water vapor can have on sensor responses, as PID sensitivity decreases dramatically with increasing RH.

Theoretically, the presence of water vapor can cause PID signals to increase or decrease with humidity: a decrease in sensor signal is explained by the absorption of UV radiation by water vapor, which reduces ionization efficiency [53], while an increase can be caused by water contamination of the sensor's electrodes, leading to an artificially high signal output due to short-circuiting [64]. Commercially-available PID sensors, including the ION Science sensors in our array, often incorporate design choices (such as a hydrophilic membrane to reduce water vapor in the detector or a "fence electrode" that can stop water-induced shorts) that can mitigate these effects [64]. As Figure 2.5 shows, we do not observe any short-circuiting at high RH. Panel (a) of Figure 2.6 shows sensor sensitivity (in mV/ppb) to isoprene and α-pinene as a function of RH. Unfortunately, the RH-induced decrease in sensitivity does not appear to be easily parameterizable. Moreover, there is a more drastic RH-induced decrease in sensitivity to isoprene than there is to α-pinene, suggesting that this effect may be compound-specific. Ultimately, while our PID sensors may advertise technologies that aim to prevent sensitivity decreases under high humidity conditions, these results suggest that PID responses can be extremely and unpredictably dependent on RH.

Figure 2.6b shows EC sensitivity (in mV/ppb) to isoprene and α-pinene as a function of RH. Unlike the PID sensors, EC sensors do not show clear trends with increasing RH for either compound.

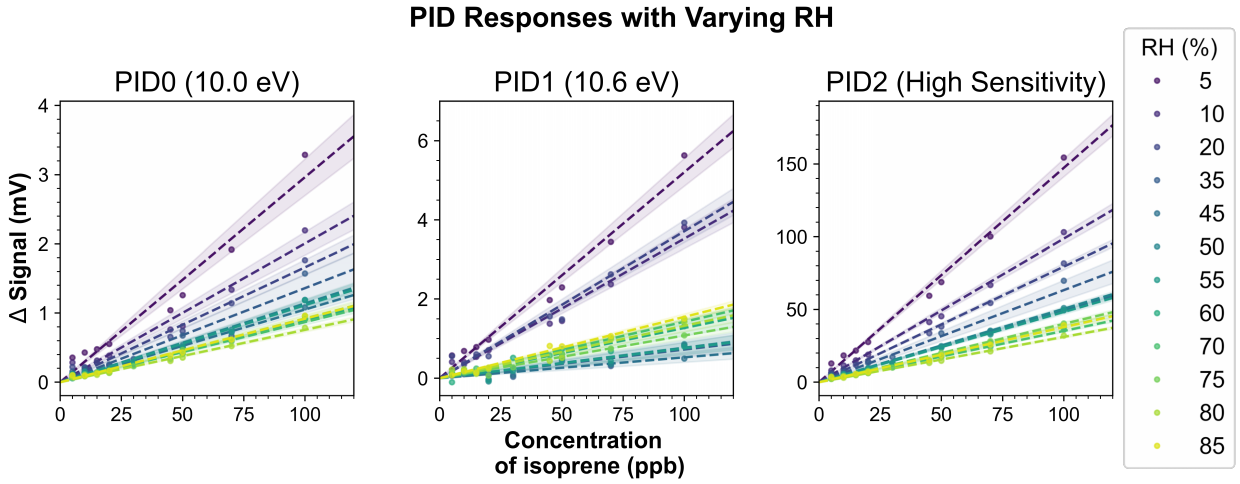

Figure 2.5 Responses of 3 PID sensors to isoprene at varying relative humidity values from 5-85%. Points denote the average of each calibration step, the dotted line showing the best least-squares fit, and the shaded area showing a 1σ confidence interval.

We initially expected our EC sensors to exhibit a more complicated relationship with humidity, as the literature shows that exposure to water vapor can affect both short-term transient responses as well as long-term sensor operating characteristics. Farquhad et al. note that short-term fluctuations of humidity are the principal cause of electronic noise in EC measurements of ambient air[65]. Exposure to water can also cause long-term changes to EC sensors, as the sensor electrolyte can increase in weight by up to 30% after being exposed to high RH, or decrease by up to 30% when dried over tens of days [62]. The RH-induced changes in the electrolyte can change the kinetics and thermodynamics of electro-oxidation, with system-dependent effects that may vary across VOCs. Due to practical considerations, we were only able to explore the short-term effects of RH on EC sensor responses. EC signals were generally much noisier than other sensors, and thus any trend in sensitivity might be partially obscured by this high noise-to-signal ratio. However, these results suggest that, without considering long-term effects, RH does not have a clear impact on EC sensitivity to these VOCs.

For MOx sensors, an increase in humidity causes decreased resistance and an increase in electron affinity. The decrease in resistance is thought to be caused by an increase in hydroxyl radicals that cover the surface of the sensor, as well as a corresponding increase in hydrogen atoms that react with surface oxygen atoms to form negative holes [66]. This effect is dependent on the sensor's operating temperature, and should be less pronounced at higher temperatures [67], [68]. Sensor response also depends on the interaction between humidity and the compounds of interest, as hydroxide ions on the sensor's surface are reactive and can oxidize the measurands, generating current.

We were particularly interested to see if changes in MOx response due to humidity could be parameterized in terms of the sensitivity A and the power law coefficient β, as prior work suggests that SnO_2 sensors measuring certain organic compounds obey $A \sim \log(\beta)$ under changing humidity [69]. Figure 2.7 shows responses of one type 1 and one type 2 MOx (both operated at 5.25V) to isoprene and α-pinene at different RH values, in terms of the response parameters A and β. For MOx type 2, the relationship $A \sim \log(\beta)$, denoted by the dotted black line, holds across rela-

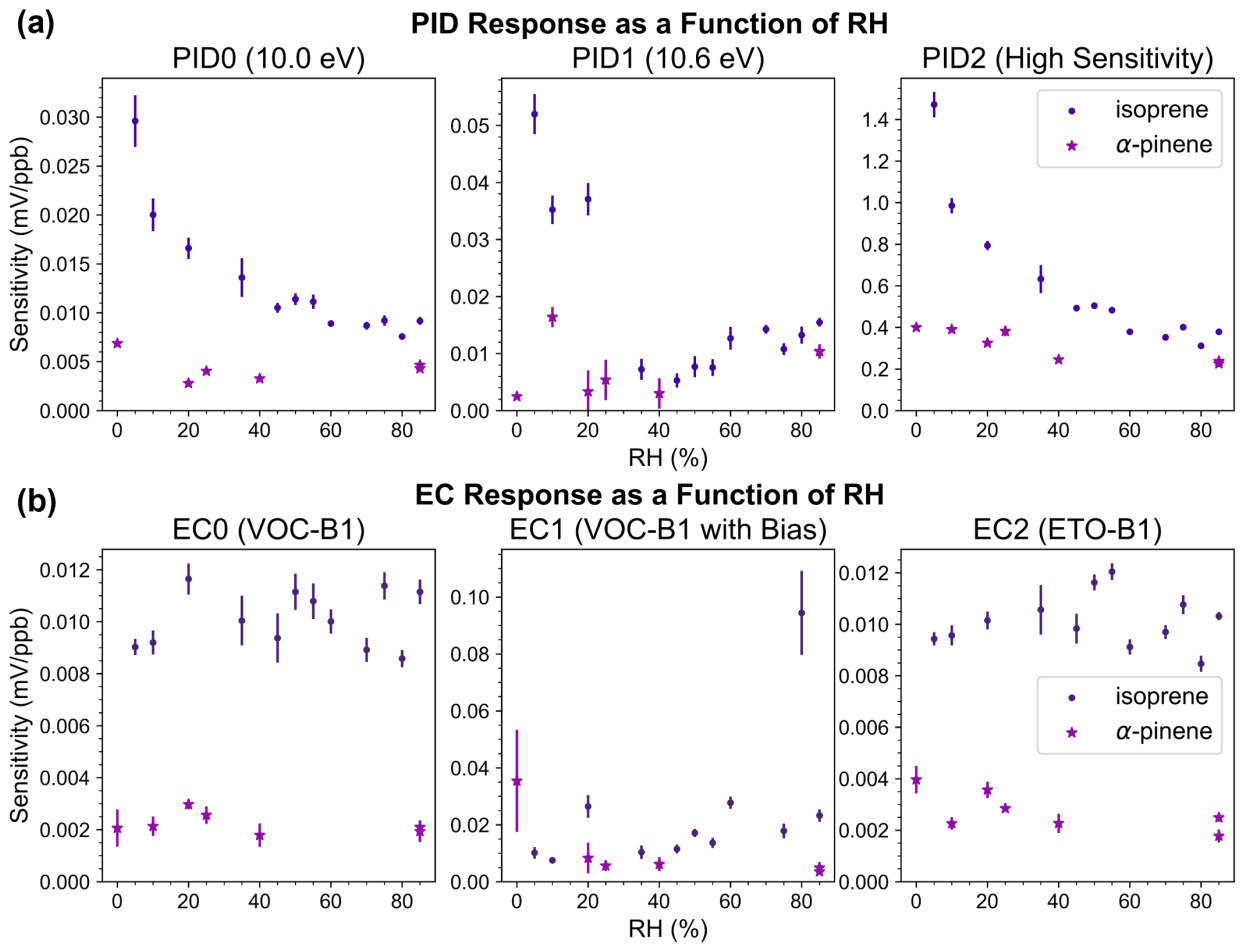

Figure 2.6 Sensitivities (in mV/ppb) of (a) 3 PID sensors and (b) 3 EC sensors to isoprene and α-pinene as a function of RH. Vertical bars denote 1σ confidence intervals.

tive humidity and compounds. These results show that in most cases an increase in RH causes an increase in the power law parameter and a proportional decrease in the log of the sensitivity. With more data, it may be possible to accurately predict the effect of changing RH on MOx type 2 responses to VOCs. However, as Chabanis et al. noted, this will be heavily compound-specific, as certain VOCs do not exhibit this response behavior [70]. Unfortunately, the relationship between fit parameters is markedly less clear for MOx type 1, and there does not appear to be a consistent relationship between increasing RH and β.

Given the compound-specific nature of PID and MOx responses to relative humidity, it would be difficult and time-consuming to parameterize the RH-dependent sensor responses for a large number of environmental VOCs in the laboratory. Our laboratory results suggest that RH extremes pose the largest challenge to the sensor array. At low values of RH (between 0 and 30%), the effect of RH on PID and MOx sensitivities is drastic and highly nonlinear, and EC sensors will experience gradual dehydration; at high RH, sensor sensitivity becomes far less dependent on RH, but we observed that degradation of PID and MOx sensors occurs after prolonged exposure. To avoid these RH-related issues, an ideal use case for this instrument would involve measurements of air either maintained at a constant and moderate RH level, or limited to a narrow range of moderate

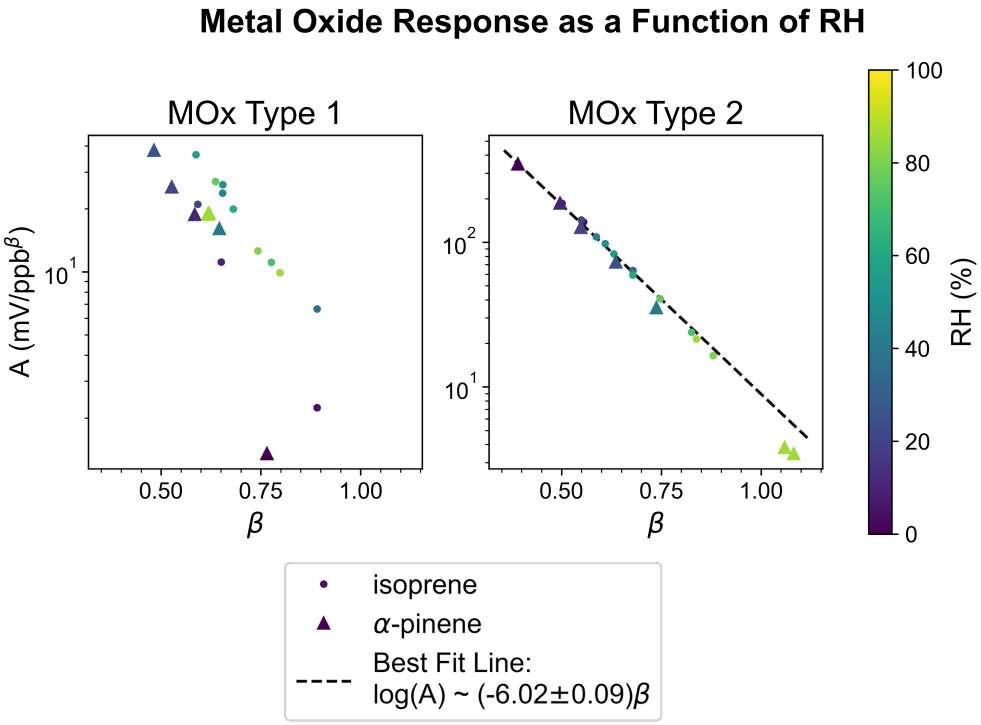

Figure 2.7 MOx responses to isoprene (circles) and α-pinene (triangles), plotted in the parameter space of dimensionless power law parameter β vs sensitivity A (mV/ppb). Points are colored by the measured RH value, and the dotted black line gives the best fit line for all data measured by the Type 2 MOx sensor.

RH values (e.g. 40 to 60 % RH). While most indoor environments will have these characteristics, typical outdoor environments will experience RH fluctuations far outside of this moderate range.

To eliminate RH-dependent changes to sensor response, practical usage of this instrument could theoretically utilize a pre-treatment technique that selectively removes water vapor from the sample stream. We originally considered the addition of a Nafion[3] dehumidifier to our low-cost instrument's inlet, as such dryer systems are commonly used to pre-treat environmental samples of ozone, CO, and CO_2 [72], [73]. Unfortunately, we found that a Nafion dehumidifer is not suitable for our VOC measurements, as it significantly depletes polar VOCs from the sample stream [74] and, without regular application of a lengthy purging process, will also cause losses and rearrangements of certain hydrocarbons [75]. There are some alternative humidity removal techniques that have been shown to remove water while preserving water-soluble sample VOCs [76]–[78], but these methods are relatively expensive and further work would be needed to identify a suitable low-cost, deployable alternative.

2.3.3 Sensor Array Responses to Basic Mixtures

Real-world environmental VOCs are most likely to be present in complicated mixtures [27], and linear additivity of LCS responses to mixture components greatly simplifies this measurement problem. Based on the theoretical principles of operation, we expect PID signals to behave in a linearly

[3] Nafion™is a copolymer of tetrafluoroethylene and fluorosulfonyl monomer [71]

additive fashion when measuring a mixture of gases [54]. Figure 2.8a shows the responses of two PID sensors to 50:50 and 75:25 binary mixtures of 1-hexene and 1-octene. For our high sensitivity PID, the mixture measurement falls exactly where one would expect, given that the individual signals for 1-hexene and 1-octene add proportionally based on their contribution to the mixture: sensor sensitivity is 0.25±4.0 × 10^{-3} mV/ppb to 1-hexene is and 1.2 × 10^{-1} ± 2.0 × 10^{-3} mV/ppb to 1-octene, and the observed sensitivity to the 50:50 mixture is 0.17 ± 2.0 × 10^{-3} mV/ppb (expected 0.18 mV/ppb), while the observed sensitivity to the 75:25 mixture is 0.20 ± 3.4 × 10^{-3} mV/ppb (expected 0.21 mV/ppb). EC sensors are also expected to behave in a linearly additive fashion [24]. While the trends in 2.8b obey linear additivity within uncertainty, high measurement noise makes the EC mixture response uncertainties much higher than those of PID sensors.

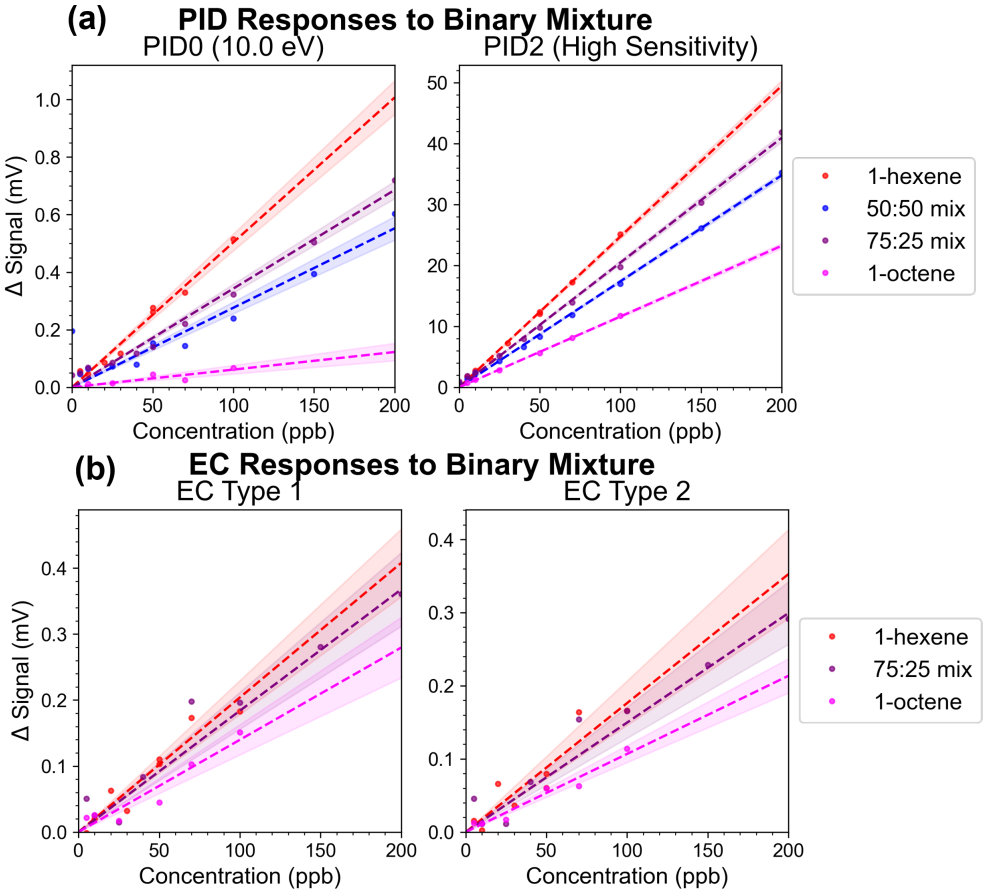

Figure 2.8 Sensor responses to 50:50 (blue) and 75:25 (dark purple) mixtures of 1-hexene and 1-octene, for (a) PID sensors and (b) EC sensors. Dotted lines indicate the best fit line for each of the points, and the shaded area indicates a 1σ) confidence interval.

In Figure 2.9, we show the response of two MOx sensors, representing both types of sensors in our array, to the binary mixture of 1-hexene and 1-octene. We have also shown the expected results of linearly adding the power law response curves with the thin dotted line: for both sensors, the observed mixture response is far below this predicted sum. This result challenges the simplifications made by early work on MOx sensors that assumed mixture components do not interact with each other, leading to a sensor response that is a linear addition of the individual power law

response curves [79].

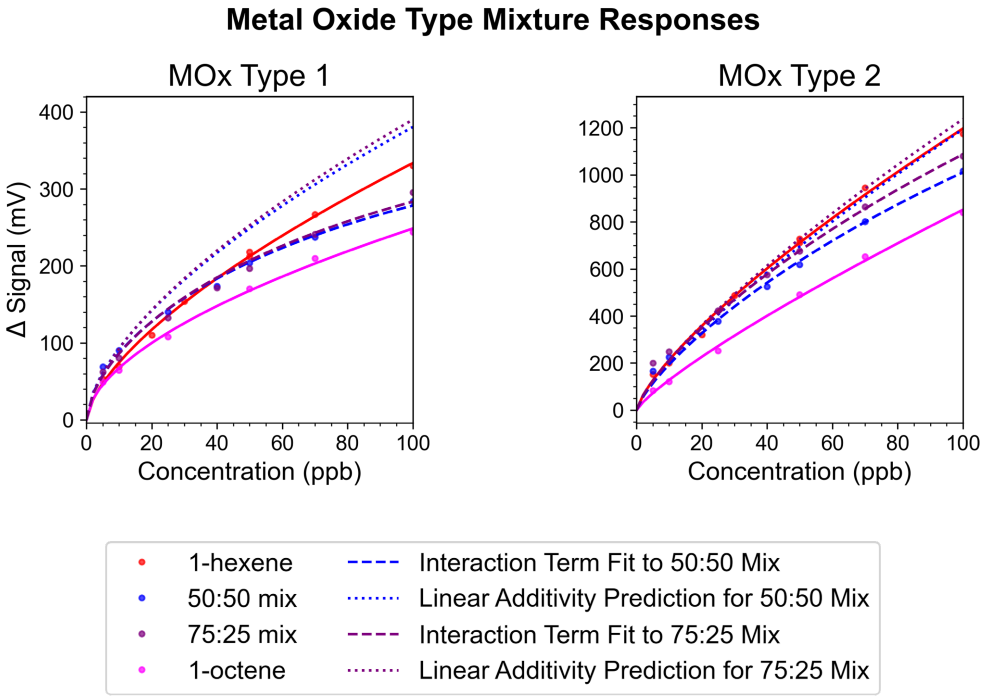

Figure 2.9 MOx responses to a 50:50 (blue) and 75:25 (dark purple) mixtures of 1-hexene (red) and 1-octene (magenta). The dotted lines represent the predicted mixture signal, assuming linear additivity. The dotted lines represent a fitted curve that includes a fitted interaction term, as described in Equation 2.1.

On the other hand, Llobet et al. assumed that this linearity assumption would hold for concentrations below 100ppm, but noticed that for higher concentrations, observed responses were lower than the predicted linear sums and were better parameterized by subtracting multiplicative interaction terms for each gas pair from the linear sum of responses [80]. In other words, they found that the response $G_{VOC1 + VOC2}$ of a MOx sensor to a binary mix of VOC_1 and VOC_2 could be represented as

$$G_{VOC_1+VOC_2} = A_1[VOC_1]^{\beta_1} + A_2[VOC_2]^{\beta_2} - A_{12}[VOC_1]^{\beta_1}[VOC_2]^{\beta_2}. \quad (2.1)$$

We applied the model proposed by Llobet to the mixture data, calculating out a coefficient A_{12} representing the interaction between 1-hexene and 1-octene. Calculated values of A_{12} were found to be consistent across the different mixture proportions: for the MOx type 1 sensor, we found $A_{12} = 2.9 \times 10^{-3} \pm 1.3 \times 10^{-4}$ for the 50:50 mixture data and $A_{12} = 3.4 \times 10^{-3} \pm 2.0 \times 10^{-4}$ for the 75:25 mixture data, while the MOx type 2 sensor's best fits had $A_{12} = 5.3 \times 10^{-4} \pm 1.3 \times 10^{-4}$ for the 50:50 mixture and $A_{12} = 5.7 \times 10^{-4} \pm 2.0 \times 10^{-4}$ for the 75:25 mixture. Our data suggests that, even though the VOC concentrations tested are relatively low, the linearity approximation has poor performance for a total mixture concentration above 10ppb. However, including a VOC-VOC interaction term does neatly explain the observed responses.

2.3.4 Characterizing VOCs with Relative Sensor Array Responses

Figure 2.10 summarizes the responses of the array to all ten compounds at 10ppb, with signals normalized to each sensor's maximum observed response at 10ppb. In general, we observe that all ten compounds are detected by some subset of the array that includes at least two distinct sensor types. Moreover, the inability of any one sensor to detect a certain compound can be a source of information about the VOC being measured, as the lack of signal still contributes to the unique "fingerprint" of each compound. The results from this figure clearly show that different sensing technologies, as well as variations within the same sensing technology, result in different responses to VOCs: in other words, each of the 12 sensors in our array appear to have their own unique sensitivities to various VOCs. It also suggests that given enough data, pattern recognition techniques could potentially be used to map a given "fingerprint" to the corresponding VOC.

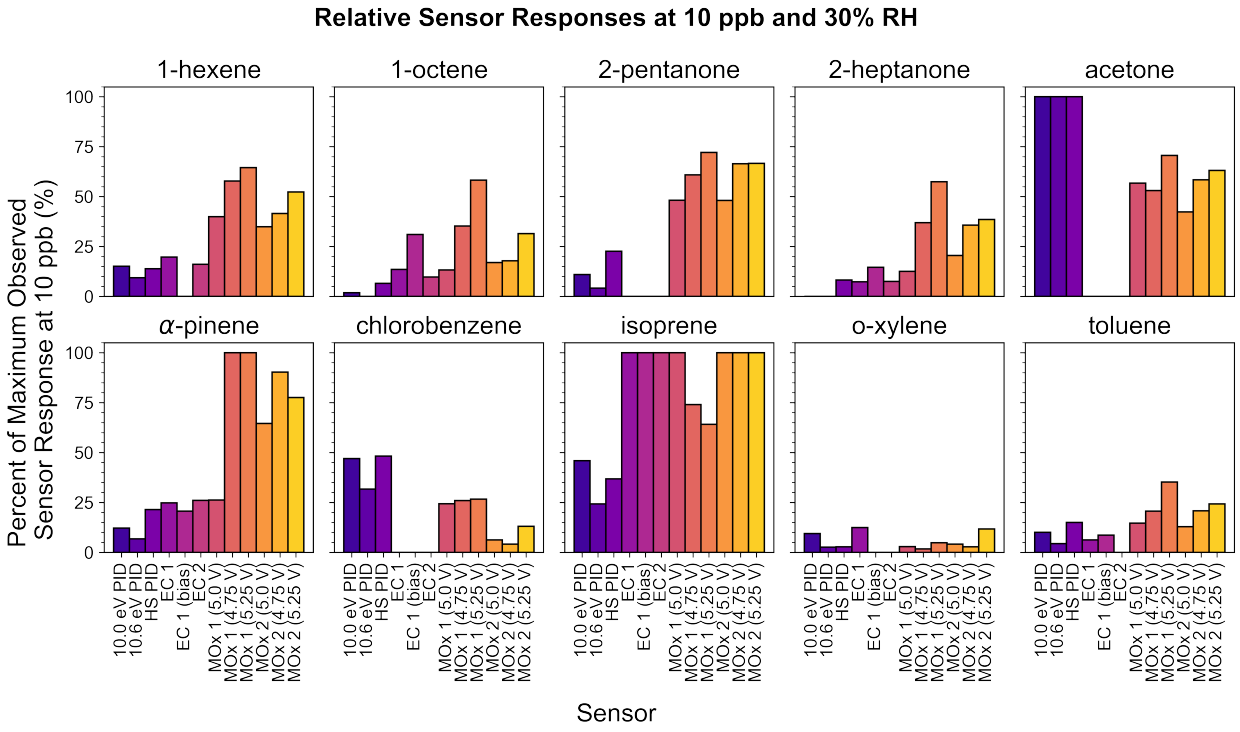

Figure 2.10 Summary of sensor responses, as a percentage of the maximum observed response for each individual sensor, to ten different compounds at 10 ppb.

2.4 Conclusions and Future Work

We designed and developed a low-cost sensor array instrument consisting of 12 total PID, MOx, and EC sensors, each with their own unique sensitivities to VOCs. We used this novel instrument to make measurements of 10 common VOCs between 5-100 ppb in a controlled laboratory environment. We also explored the effects of changing RH: PID sensors saw a consistent but VOC-specific decrease in sensitivity with increasing RH, EC sensor sensitivities did not change much or in a predictable manner, and one type of MOx sensor saw changes in its signal fit parameters that were consistent across different RH values, while the other saw significant and unpredictable changes in signal due to RH. For a simple binary mixture, we observed that PIDs and EC signals obeyed linear additivity, while MOx sensors did not. These results are largely consistent with our expectations based on prior studies and physical sensing principles, and figure 2.10 highlights the potential of such a sensor array in making measurements of VOCs: the entire array of responses clearly gives us unique information about the VOC being measured.

However, our work on RH and mixture responses highlights an inconvenient truth about transferring laboratory results to the practical usage of this instrument. In a real environment where our sensor array might be used, RH will likely fluctuate by large amounts, and the VOCs being sampled will almost certainly be present in complicated mixtures. Our results demonstrate that the effects of RH and mixtures are significant and are very difficult to parameterize. The effect of water vapor appears to depend strongly on the specific VOC being measured, and characterizing this effect for every possible environmental VOC would be unrealistic. Moreover, while we observed that PIDs and ECs have linear additive responses to binary mixtures, we also saw that MOx responses are heavily dependent on a fitted interaction term unique to the two gases in the mixture. Accurately predicting MOx responses to mixtures of VOCs would thus involve determining interaction terms for every possible pairing of VOCs expected to be seen by the array, a task that is unlikely to be accomplished quickly or inexpensively.

Our results are far from a complete laboratory characterization of this sensor array: notably, we have not attempted to characterize the transient behavior of these sensors (e.g. signal response and decay times), nor have we conducted laboratory experiments to quantify baseline drift due to ambient temperature changes or sensor age. Future work could certainly attempt to characterize these unaddressed aspects of sensor responses. It could also attempt to fill in the blanks of this study by testing more VOCs than the 10 explored, determining sensitivity decreases as a function of RH and the VOC in question, or exhaustively determining MOx interaction terms for many different gas pairings.

However, we believe that such work is simply not scalable: the real atmosphere contains far too many VOCs, with too many RH conditions and possible VOC mixture compositions, for such prescriptive lab characterization to be feasible. Consider the hypothetical case of fully characterizing sensor responses to 30 important environmental VOCs, a conservative number given the complexity of ambient air. Sensor characterization at three possible RH values (low, medium, high) would already require a minimum of 90 laboratory experiments, and if MOx interaction terms were desired the number of experiments would balloon due to the 435 possible unique pairings. It may be possible for future laboratory work to avoid these issues by designing calibration processes that are more tailored toward environmental applications: for example, future calibrations could be focused on characterizing a sensor array's responses to real atmospheric mixtures representing emissions

from VOC sources like biomass burning and traffic.

Given the certain challenges in applying laboratory characterization to real-world measurements, we recommend that future work on low-cost VOC arrays focus on characterizing sensor responses to real-world VOC mixtures rather than to exhaustive, ground-up laboratory experiments. This could be achieved through an application-focused calibration of a low-cost VOC array based on real-world mixtures, or through the novel application of analysis techniques to directly interpret in-field LCS measurements, an example of which we will discuss in Chapter 3.

Chapter 3

Inferring Indoor Pollution Sources from a Low-Cost VOC Sensor Array

3.1 Introduction

Indoor air quality and chemistry is of particular interest to human health, as humans generally spend the majority of their time indoors [8] and exposure to household air pollution is among the top ten risk factors for disease worldwide [11]. VOCs are present indoors at concentrations that are generally much higher than typical outdoor concentrations [45], and contribute to multiple adverse health effects such as childhood asthma and allergies [9]. Indoor VOCs come from a variety of sources: a substantial portion of these compounds are emitted continuously from the building and its static contents [12], but VOCs are also released during common household activities such as cooking or application of personal care products [13] and are even emitted in significant quantities from human skin, clothing, and breath [81]. Real-time measurements of indoor VOCs can be made using proton transfer reaction mass spectrometry (PTR-MS) or chemical ionization mass spectrometry (CIMS) [32], [33], both of which generally measure at $<$1Hz and represent a vast improvement over the coarse, several-hour time resolution of traditional sorbent tube gas-chromatography (GC) measurements [34]. However, these state-of-the-art techniques also have much higher material cost and operation requirements, making it difficult to achieve spatially-distributed measurements with high time resolution that could be of interest in an indoor environment where occupant exposure to indoor emissions can be heavily influenced by location within a room or building[82].

Some past studies have recognized that low-cost VOC sensors could be useful in making spatiotemporally distributed measurements of indoor air, especially given that the higher indoor concentrations of VOCs partially compensate for the sensors' relatively higher limits of detection. One example is the development of a metal oxide microarray for indoor air quality monitoring by Arnold et al [48]. In the laboratory, the microarray was able to distinguish typical fire emission gases (benzene, formaldehyde, and acrolein) at high concentrations; in the field, the array's responses to two real-world pollution sources (smoldering cables and an occupied meeting room) could be clearly distinguished from the its responses to isopropanol-soaked tissues, though these sources represent extreme concentrations of VOCs that limit the transferability of study results to typical indoor air. Another example is the array of four different metal oxide sensors and an RH/T sensor utilized by Zhang et al. to detect and classify six common indoor air contaminants at $>$ 1 ppm

concentration (including benzene, toluene, and formaldehyde) [46]. While this study successfully distinguished sensor array responses to each contaminant via a non-linear classification algorithm applied to laboratory data, it did not attempt to extend this approach to real indoor air. Similarly, Wolfrum et al. also attempted to develop and characterize a metal oxide sensor array intended for building environmental monitoring [47]. Though the researchers were able to obtain calibrations for several common VOCs in a relevant concentration range, they recognized the challenges of transferring their laboratory results to an actual indoor environment—namely, that spatially-distributed measurements would require individual calibrations for every single sensor array, and that this calibration process would be made even more infeasible by the large number of important VOCs in the indoor environment to calibrate for. In summary, while there have been several attempts to develop a low-cost indoor VOC monitor using laboratory calibrations, to our knowledge there have been no attempts to measure realistic and common indoor VOC pollution sources in the field with a low-cost sensor array.

Given the challenges associated with calibrating low-cost sensors for individual compounds, some studies take a different focus: sensor data is used to reveal underlying factors that control pollution e.g. identifying sources of VOCs. This is not a wholly new concept in sensor array research: pattern recognition techniques such as linear discriminant analysis or principal component analysis have long been used in "electronic nose" research to identify and classify sensor array responses to odors [83]. Whether similar techniques would work for low-cost measurements of ppb-level environmental VOCs has yet to be truly tested. Results from the simplest case of a two-sensor array certainly show promise: Collier-Oxandale et al. compared measurements at an outdoor site near Denver made by two distinct metal oxide sensors, and noticed that differences in signal between the two sensors could be leveraged for information on chemical sources—namely, that the ratio of the two metal oxide responses, due to the sensors' varying sensitivities to toluene and benzene, could be used to separate measurements of traffic from measurements of oil-and-gas emissions [44]. It seems feasible, then, that data from a larger sensor array could provide even more information on the underlying sources of VOCs and other pollutants.

Here, we describe measurements, made by the novel low-cost VOC sensor array described in Chapter 2, of indoor air subject to experimental perturbations representing realistic sources of VOCs, such as cooking, pesticide application, and wildfire smoke intrusion. We show that even without calibration, the array of responses can provide information about the composition of VOCs in indoor air. Moreover, through application of non-negative matrix factorization to both our low-cost sensor and reference datasets, we show that the differences between sensor responses in our array can be leveraged to give both quantitative and qualitative information on multipollutant sources in indoor air.

3.2 Methods

3.2.1 Indoor Air Measurements

Measurements were taken as part of the Chemical Assessment of Surface and Air campaign (CASA), a collaborative indoor field experiment that ran from late February to early April of 2022 at the Net Zero Residential Test Facility (NZRTF), a model two-story residential home in Gaithersburg, Maryland. Table 3.1 details some of the experiments performed during CASA to perturb indoor air

quality and chemistry.

Table 3.1 Summary of CASA activities.

Activity Name	Description
Chemical "Cocktail"	Known VOCs (o-xylene, chlorobenzene, α-pinene, 1-hexene, 1-octene, acetone, 2-pentanone, 2-heptanone, toluene-d8, furfural) were heated together in a flask, submerged in a heated (85-100 °C) water bath, and connected to Teflon tubing carrying zero air flow into the test house living room from the porch. Cocktail injections varied in their inclusion of furfural, as well as the volume of flask compounds—each VOC was included in the mixture at the same individual volume of either 0.05, 0.5, or 1 mL.
Pesticide	20-40 g of a common pesticide was sprayed by a participant around the first floor of the test house for about one minute, with a box fan turned on for the duration of the product application.
Cooking	Various foodstuffs (such as chicken tenders, popcorn, balsamic vinegar) were prepared and cooked via air-frying, pan-frying, or microwave in the test house kitchen.
Acid/Base	Either CO_2 or NH_3 was injected to the test house from the attic continuously for about 30 minutes, until reaching a high but realistic concentration (5000 ppm CO_2, 500 ppb NH_3).
Fresh smoke	0.50g of woodchips were loaded into the basket of a portable cocktail smoker, and a participant entered the first floor of the test house to deploy the smoke in front of a box fan for about one minute.
Aged smoke	A 1m^3 Teflon chamber outside of the test house was gradually filled with smoke and ozone (with [O_3] maximum of about 10 ppm in the chamber), then left to marinate for 1 hour. The chamber was then connected to an inlet on the side of the house, and a zero-air generator was used to purge the chamber contents into the ventilated first floor for about 30 minutes.
Surface/air cleaning	For surface cleaning, participants mopped the floors using a solution of commercial product, vacuumed the floors, and dusted other surfaces. For air cleaning, several different commercially available air cleaners were each operated remotely for 30 minutes at a time.
Tracer	Every night at 2:00 a.m. EST, a pulse of sulfur hexafluoride was automatically injected into the first floor of the test house. The maximum concentration in the living room was normally about 10 ppm.
Environmental Parameters	For at least one full day a week, relative humidity at the test house was raised using an in-duct humidifier system to a maximum of about 75%. During temperature ramp experiments, the temperature of the unoccupied house was ramped from 20 to 30 °C and held for 2-3 hours, then allowed to passively cool to the house's setpoint. Occasionally, other experiments were run under high RH conditions (e.g. cooking, cocktail, fresh smoke).

Reference measurements were made by a GC-Vocus PTR-ToF-MS (Gas Chromatograph coupled to a Vocus-Proton Transfer Reaction-Time of Flight-Mass Spectrometer), which we will refer to simply as the GC-Vocus. PTR-ToF-MS makes measurements by ionizing VOCs via proton transfer from hydronium ions, which then allows for the ions to be detected based on a time of flight that is mass-dependent [84]. Using a fast GC inlet in tandem with PTR-ToF-MS allows for some molecular identification of multiple isomers, cluster ions, or fragmentation products that share the same mass [85]. We will primarily be using uncalibrated data in units of ion counts per second (cps), which have been zero-subtracted and corrected for the variation in the instrument sensitivity during the full campaign. Vocus-PTR-MS sensitivities vary across compounds and are generally in the range of 1000-9000 cps ppb^{-1} [40], a range that is narrow enough for our uncalibrated reference data to serve as a rough proxy for actual VOC concentrations. A few confident calibrations were obtained

for this dataset, including sensitivities of 3800 ± 260 cps ppb^{-1} for furan and 3000 ± 390 cps ppb^{-1} for toluene that are within the estimated sensitivity range. Assuming an average sensitivity of 3500 cps ppb^{-1}, we estimate that the maximum total VOC concentration observed in the test house living room (achieved during an early "chemical cocktail" experiment) could exceed 1 ppm, with a baseline concentration of around 150 ppb.

We made measurements using the same low-cost VOC instrument described in detail in Chapter 2. Briefly, this is an array of 12 low-cost sensors, representing three fundamentally different sensing technologies: photo-ionization detection (PID, 3 sensors total), electrochemical (EC, 3 sensors) sensing, and metal oxide (MOx, 6 sensors) sensing. We were able to vary operational or physical parameters between sensors with the same sensing technology such that each of the 12 sensors has its own distinct set of sensitivities to various VOCs. We deployed three of these low-cost instruments in the test house but here our analysis focues primarily on results from a single instrument that was co-located with most of the gas- and particle- phase reference instruments in the living room area of NZERTF. We also made co-located low-cost measurements of CO and particulate matter (PM) using a QuantAQ Modulair air quality monitor [86]. The Modulair uses an Alphasense CO-B1 sensor to make measurements of CO and uses two different low-cost optical particle sensors (Alphasense OPC-N3 optical particle counter and Plantower PMS nephelometer) to make size-resolved particle measurements. The measurements made by LCS that are used in this analysis, along with the abbreviations we will be using to refer to them, are summarized in Table 3.2.

Table 3.2 Summary of LCS measurements used in NMF analysis.

Name	Description	Name	Description
PID0	ION Science MiniPID 2 10.0eV	MOX3	Figaro TGS2602 or "Type 2" MOx (5.0 V supply voltage)
PID1	ION Science MiniPID 2 10.6 eV	MOX4	Figaro TGS2602 (4.75 V supply voltage)
EC0	Alphasense VOC-B4 VOC sensor (no bias)	bin0	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.38-0.46 µm
EC1	Alphasense VOC-B4 VOC sensor (positive bias)	bin1	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.46-0.66 µm
EC2	Alphasense ETO-B1 VOC sensor	bin1	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.66-1.0 µm
MOX0	Figaro TGS2600 or "Type 1" MOx (5.0 V supply voltage)	Dewpoint	Dewpoint as calculated from RH/T measurements (Sensiron SHT25)
MOX1	Figaro TGS2600 (4.75 V supply voltage)	CO	Alphasense CO-B1 sensor)
MOX2	Figaro TGS2600 (5.25 V supply voltage)		

3.2.2 Non-Negative Matrix Factorization

Here, we describe the application of non-negative matrix factorization (NMF) to both our reference and low-cost datasets. NMF is an unsupervised dimensionality-reduction technique, where a matrix

K with size $m \times n$ (representing in our case the n species observed at m different time points) is separated into the product of two lower-dimensional matrices and H, with sizes $m \times k$ and $k \times n$ [87]. We chose to use an NMF algorithm, implemented in the scikit-learn python library, that minimizes the Frobenius norm between K and its approximation W × H via use of a coordinate descent solver [88]. Because of the non-negativity constraint on both the original matrix and the lower-dimensional approximation, the solution is physically interpretable: for example, if the i^{th} column in K represents the time series of a single pollutant, then the columns of W represents the time series for k sources that each contribute fractional amounts of signal represented by the i^{th} row of H. NMF has previously been shown to yield physically-meaningful results for VOC measurements made by a ToF-chemical ionization mass spectrometer [89]. It has also been used on low-cost sensor array data, such as when Hagan et al. [22] applied NMF to a low-cost sensor array measurements of inorganic criteria pollutants and particulate matter and found that the resulting factors were representative of known aerosol pollution types in Delhi, India.

One of the key issues with dimensional reduction techniques such as NMF is the correct determination of the rank k, which greatly impacts the interpretability of results. There is no universally-accepted approach to this question, and some applications of factor analysis for environmental data even sidestep the mathematical aspect of this problem by prioritizing the physical interpretability of their results at different values of k [89], [90]. One example of rank determination for environmental NMF solutions is found in Hagan et al. [91], where a bi-cross-validation technique was applied to LCS data: this method repeatedly trains an NMF model on a block of randomly selected samples and features and then evaluates them on a non-overlapping block of samples and features [92]. Another possible approach is imputation cross validation: in these schemes, a small fraction of values in K are "masked" during the factorization process, and the error between the "masked values" and imputed ones is calculated after model training [93]. Here, we choose to use a recently developed imputation technique (CV2K) which was shown to outperform bi-cross-validation techniques in choosing the correct rank for synthetic datasets [94].

We chose to limit our analysis of the reference dataset to only the most important VOC signals observed over the course of the campaign. We chose these compounds through a simple filtering process. First, we screened for duplicate signals, as VOCs can fragment within the Vocus-PTR-MS or be detected as water adduct ions [95], [96]. This was achieved by identifying groups of ion signals with Pearson $r^2 > 0.99$ and removing all but the signal corresponding to the heaviest compound. Then, we formulated a cumulative list of all VOCs that, for any time point in the minute-averaged dataset of the 26 campaign days measured by the GC-Vocus, appeared in the top 20 most intense signals. This filtering process allowed us to reduce the reference dataset by an order of magnitude, narrowing our focus from over 600 ion signals to a subset of 65. Reference data was then averaged over ten minutes and each individual time series was normalized via min-max scaling to the range [0,1], resulting in a $65 \times m$ input matrix, where $m = 3691$ is the number of observations. We also conducted the subsequent analysis without normalization (see Appendix section 3.A.4) but found that the scaled version yields a far more meaningful representation of VOC sources: in the solution without normalization, the disparate relative magnitudes of various CASA activities led to drastic over- or under- representation of known sources (e.g. several factors representing different "chemical cocktail" emission patterns were resolved, but there were none representing wildfire smoke intrusion).

We also applied NMF to the LCS signals described in Table 3.2 to "raw" output sensor voltages, with each individual time series averaged to the same time intervals as the reference dataset and

normalized to the range [0,1], that are assumed to be linearly proportional to pollutant concentration. As shown in the results from Chapter 2, this is a good assumption for many of our low-cost VOC sensors but may be somewhat inaccurate for metal oxide sensors, which have a power-law dependence on concentration and respond to VOC mixtures in a nonlinear manner. However, MOx sensor responses are most nonlinear at small VOC concentrations—and we posit that the indoor total VOC baseline concentrations during CASA were so high (tens to hundreds of ppb) during CASA that MOx responses to any chemical perturbations should be approximately linear. We also know from subsection 2.3.2 that our sensor signals to various VOCs may depend on humidity, with the largest decreases in sensitivity observed from 0 to 40% RH (see Figures 2.5, 2.6, and 2.7). During CASA, the baseline RH was $\sim$ 30%, while the maximum observed RH was $\sim$ 60%. This observed range is narrow enough that RH-induced changes of sensor VOC sensitivities should be relatively minor. Nonetheless, sensor baselines may still be affected by environmental RH and temperature, but this effect should be mitigated by the inclusion of environmental dewpoint in our analysis inputs as we discuss later. Despite similar uncertainties associated with their LCS measurements, Hagan et al. showed that the use of uncorrected or “raw” LCS data can still yield meaningful results about outdoor PM pollution components [91]. Similarly, the use of "raw" data in this analysis allows us to demonstrate that LCS can be used to infer indoor pollution sources even without a comprehensive characterization of sensor responses to VOCs, gas mixtures, and interfering factors.

3.3 Results and Discussion

3.3.1 Source Apportionment of Reference VOC Dataset

We applied NMF to our filtered and scaled GC-Vocus dataset, using an imputation cross-validation technique to arrive at a 15-factor solution. More details about the cross-validation, as well as the plot of imputation errors, can be found in the Appendix of this chapter (3.A.1). It should be noted that this number of factors is higher than typical numbers obtained from factor analysis studies of outdoor environmental VOC data, which generally arrive at solutions of seven or fewer factors [37]. This can be explained by the unique conditions of our indoor measurements compared to typical measurements of outdoor sources, as our reference measurements are made in proximity to a number of very strong sources (shown in Table 3.1). As shown below, most of these factors are clearly associated with CASA experimental activities, suggesting that the NMF solution’s high number of factors accurately reflects the complexity of the indoor environment. We were able to qualitatively identify these source profiles using our pre-existing knowledge of activities that introduced VOCs into the test house, but we also developed a quantitative approach for factor identification based on the timestamps of known CASA events. The relative magnitudes of the source profiles, in units of cps, were determined by fitting a multiple linear regression (MLR) model of factors to the total summed signal of relevant ions. The quantitative factor identification is described in Appendix section 3.A.2, while the determination of source profile magnitude in cps is detailed in Appendix section 3.A.3.

Figure 3.1 shows NMF results for the reference dataset plotted as time series across the entire CASA experiment, with colored dots indicating the primary activities for each day of experiments as detailed in Table 3.1. As each of the original VOC ion signals can be represented as the sum of contributions from factors, this information can be used to calculate rough compositional data for

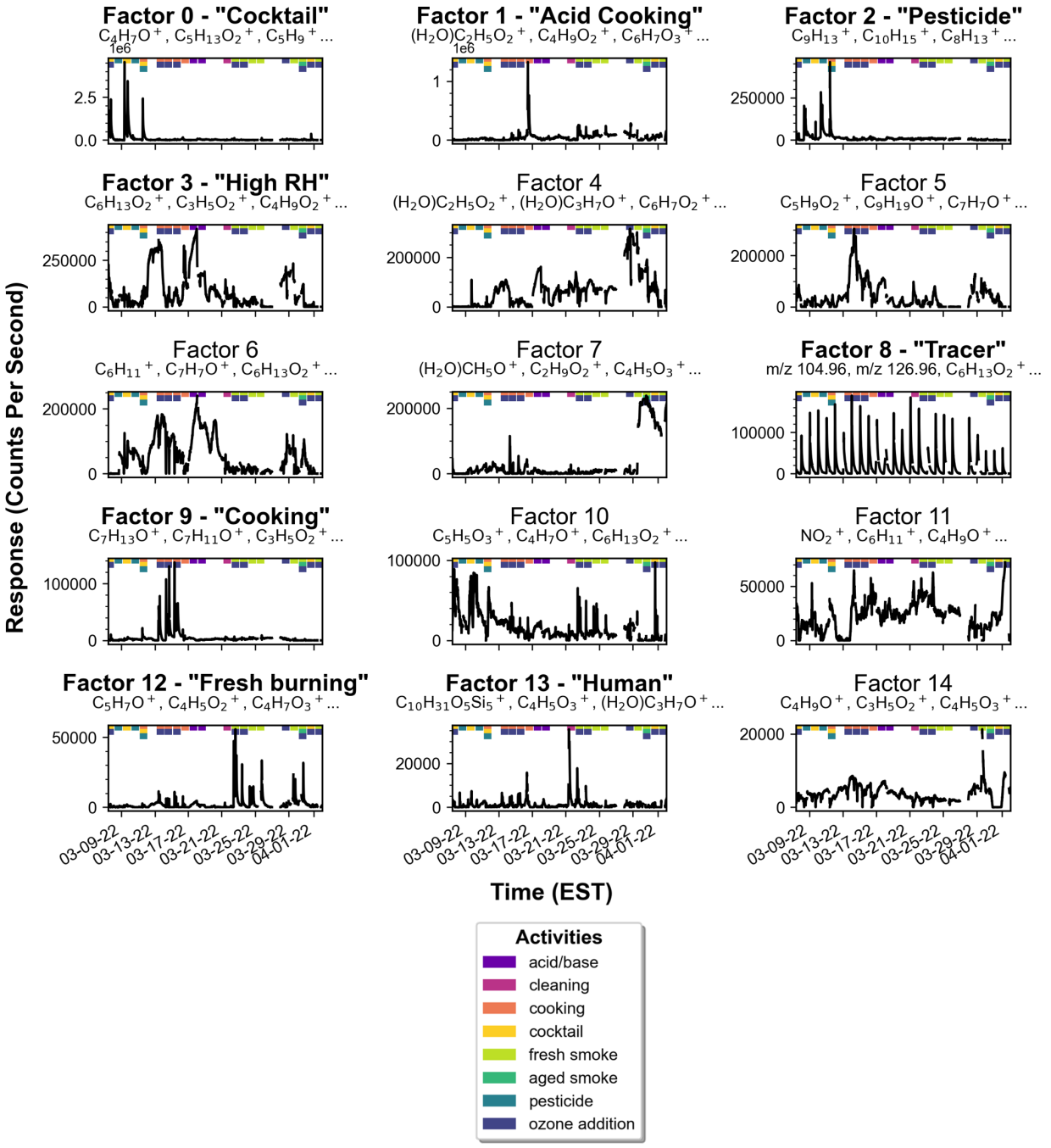

Figure 3.1 15-factor NMF solution for scaled input reference VOC data. Colored dots indicate the primary activity for each day of experiments, as detailed in Table **tab:2**. The subtitles of each panel also list the three ions that are most strongly associated with each factor, which means that the factor explains more of the variation in these ions' signals than it does of variations in other signals.

each factor. Appendix Section 3.A.2 elaborates a bit more on the assumptions made to generate the chemical compositions of each factor as shown in Figure 3.2. We see that each source profile, in addition to having a distinct time series, also has a unique chemical composition that reflects the

diversity of indoor pollution sources.

To quantitatively associate factors with a certain pollution source, we multiplied binary activity series representing known activities by the identified peaks of each factor, then tabulated the sums of these products as shown in Appendix Figure 3.A.2. Factors that we were able to associate with a certain pollution source are labelled in Figure 3.1: we see that there are factors that clearly represent source profiles for the chemical cocktail and pesticide experiments, cooking activities, wildfire smoke experiments, the SF_6 tracer injections, and VOC emissions caused by human occupancy.

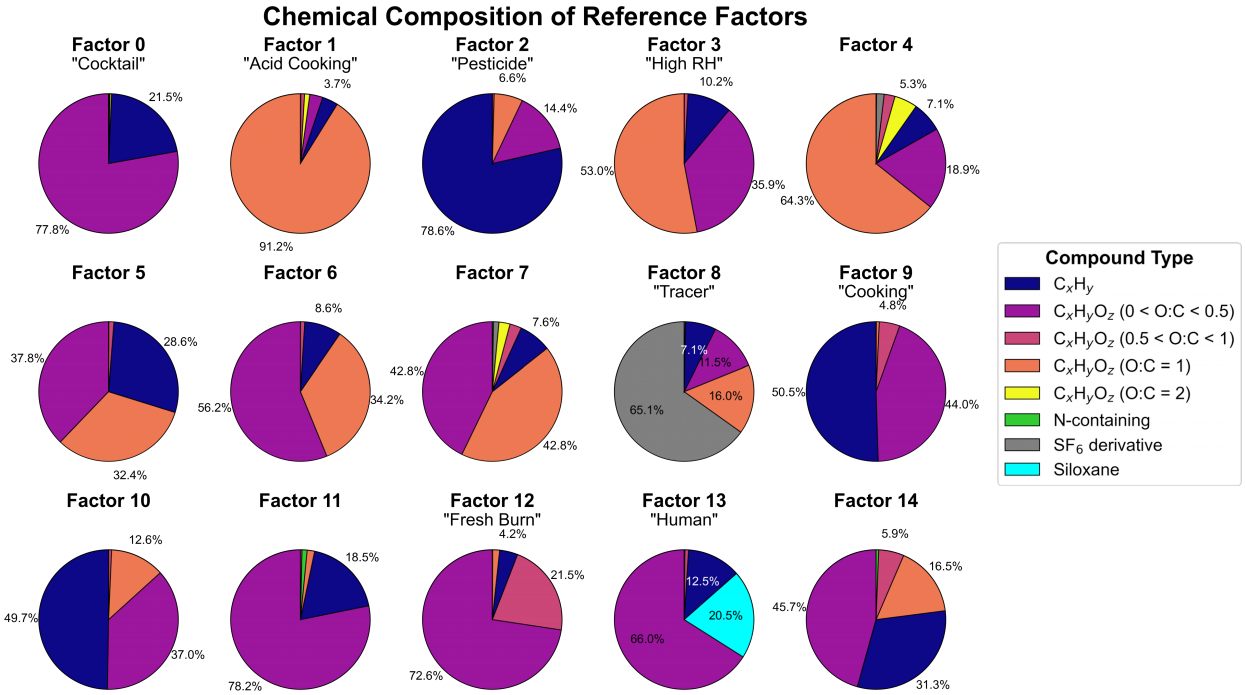

Figure 3.2 Fractional compositions of each reference NMF factor, expressed in percentage of source profile signal. Compounds are lumped into different chemical categories based on their molecular formula.

We observe that some of these factors (3, 4, 5, 6 and 14) are not associated with known activities in the house but have moderate correlations with periods of either high humidity or high temperature (see Appendix figure 3.A.9). It is possible that one or more of these factors are influenced by an RH dependence of Vocus-PTR-MS sensitivities to certain compounds. For example, Li et al. showed that the presence of humidity has nearly no effect on Vocus-PTR-MS sensitivity to terpenoids, but slightly decreases instrument sensitivity to aromatic hydrocarbons by about 1% per 10% increase in RH and significantly increases sensitivity to long-chain aldehydes to a maximum of 4% increase for every 10% increase in RH [96]. Over the course of CASA, RH ranged from a baseline of about 30% to a maximum of about 60%, suggesting a potential 12% increase in Vocus-PTR-MS sensitivities to long-chain aldehydes that could be contributing to the trend seen in Factor 5, which is strongly associated with an ion ($C_9H_{19}O^+$) that could be nonanal.

However, it is more likely that these factors reflect actual emission trends of indoor VOCs, as both field studies [97] and laboratory experiments [98], [99] have shown that concentrations of certain VOCs are strongly dependent on environmental parameters. This effect is explained by competitive adsorption: when RH increases, water competes with VOCs that are adsorbed onto

indoor surfaces, causing certain VOCs to be released into the air; when temperature increases, adsorbed VOCs tend to evaporate faster than adsorbed water does [97]. For example, acetic acid is emitted from wood and its indoor concentrations are known to depend heavily on RH and temperature[100]. Our NMF results demonstrate this clearly: acetic acid, detected as $(H_2O)C_2H_5O_2^+$, is strongly associated with both the "acid cooking" factor and the RH/temperature-dependent factor 4, indicating the contributions of both human activities and indoor surfaces to its indoor concentration.

3.3.2 Application of NMF to Low-Cost VOC Measurements

We applied NMF to the scaled LCS measurements described in Table 3.2 and shown in Appendix figure 3.A.7, using the same imputation cross-validation technique applied to the NMF decomposition of the reference data; cross-validation results are summarized in Appendix section 3.A.6. Here, we describe a 10-factor solution, with the time series of each factor displayed in Figure 3.3. We also show the percentage of each of the original LCS signals that is explained by each factor in Figure 3.4. In Figure 3.3, a handful of the factor time series appear to resemble some of the VOC source profiles seen in Figure 3.1. 3.4 shows that every LCS factor is meaningfully associated with multiple low-cost VOC sensor signals. For example, LCS factor 1 correlates well with the wildfire smoke experiments and is most strongly associated with the observed PM signal but also helps to explain small fractions of signal from PID, EC, and MOx sensors.

To explore the relationship between these LCS-derived factors and indoor pollution sources, we correlate each factor's time series to the reference factors' time series obtained earlier. The results of this correlation analysis are shown in Figure 3.5. There are a few pollution sources whose time profiles have high correlation with LCS-derived factors. The LCS-derived factors clearly capture the chemical cocktail, fresh burning, and pesticide emissions. We see from Figure 3.4 that each of these sources is associated with a different subset of sensors: LCS factor 0, which corresponds with the chemical cocktail emissions, is primarily associated with PID sensors; LCS factor 1, which captures the fresh burning factor, is primarily associated with low-cost PM measurements and a mix of small contributions from all of the VOC sensors; LCS factor 2, which generally follows the pesticide emission profile, is primarily associated with MOx sensor signals. We also notice a moderate correlation between LCS factor 6 and unidentified reference profile 10, which our quantitative analysis showed to be mildly related to smoke activities but could represent a real emission profile that can't be identified based on known CASA activities. We also applied the quantitative factor identification technique described in Appendix section 3.A.2, with results shown in Figure 3.A.11, and found that this alternative approach accurately identifies these event-based, LCS-derived source profiles in the absence of reference VOC data.

LCS factors 3, 7, and 8 also have moderate correlation with reference factors 6, 3 and 4 respectively, which, as we discussed earlier, are likely be emission profiles of VOCs from indoor surfaces caused by higher RH/T conditions. It is possible that these LCS factors simply represent RH-dependent baseline changes, in which case we would expect them to explain a large fraction of the dewpoint trend and be primarily associated with sensors that have a strong baseline dependence on dewpoint (e.g. MOx Type 1) [60]. Figure 3.4 shows that factor 7 partially fits this profile, but factors 3 and 8 explain only a small portion of the variation in dewpoint. Moreover, some of the sensors associated with these factors (primarily MOx type 2 and the unbiased EC sensors) are not expected to have strong baseline dependencies on RH [61], [65]. Thus, these LCS factors could be meaningfully correlated to reference factors that represent indoor surface emissions.

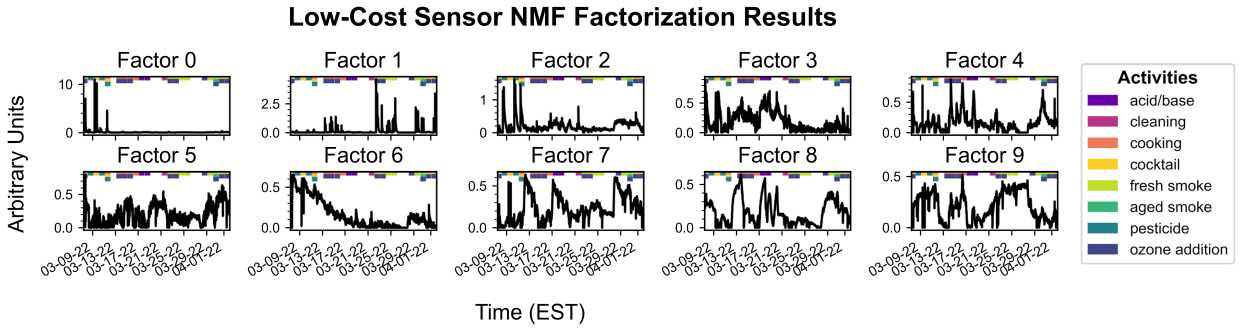

Figure 3.3 10-factor NMF solution for the low-cost sensor array data, shown as time series over the course of the CASA experiment in arbitrary units.

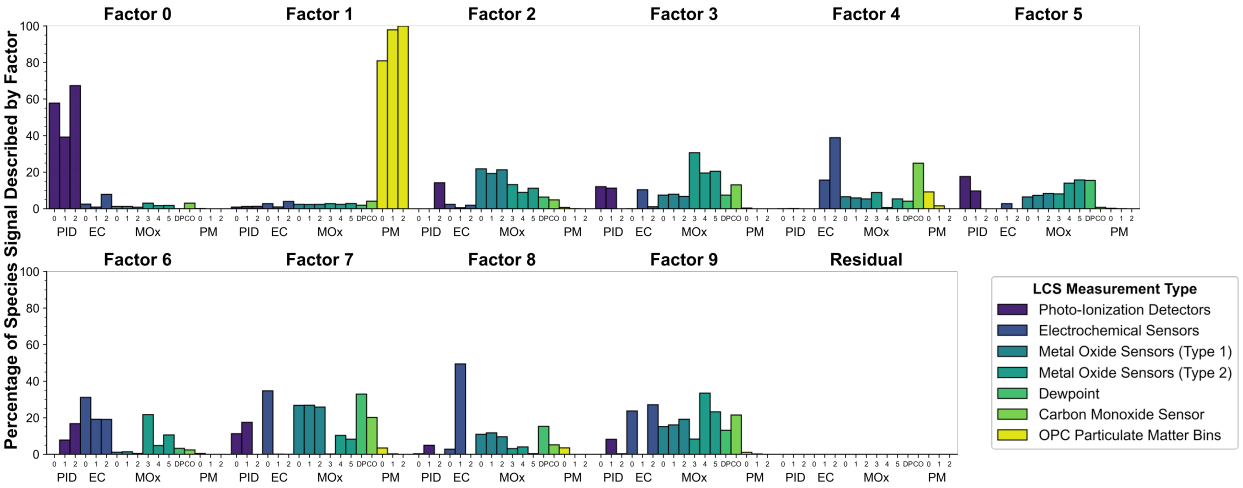

Figure 3.4 Fraction of each input low-cost sensor measurement associated with a given LCS-derived factor. For example, Factor 0 explains 45-50% of the signal from PID sensors, along with small fractions of observed responses of other low-cost VOC sensors; the rest of the observed PID signal is spread out amongst the other factors.

The compositional data of each source can also give some information about the compounds that each sensor is the most sensitive to. For example, the "chemical cocktail" signal is primarily made up of $C_xH_yO_z$ (O:C < 0.5) compounds with a smaller fraction of hydrocarbons (C_xH_y), while the pesticide signal is primarily made up of hydrocarbons with a smaller fraction of $C_xH_yO_z$ (O:C < 0.5) compounds. The LCS factor that is correlated to the "chemical cocktail" profile is primarily associated with PID sensors, with small contributions from MOx sensors; on the other hand, the LCS factor that is correlated to the pesticide emissions is primarily associated with MOx sensors, with small contributions from PID sensors. From this observation, we may conclude that PIDs are more sensitive to $C_xH_yO_z$ (O:C < 0.5) compounds, while MOx sensors are more sensitive to hydrocarbons. To aid in these associations, Figure 3.A.12 in the Appendix shows the relevant subplots in Figure 3.4 compared with reference factor compositions for LCS factors that showed at least moderate ($r > 0.6$) correlation with a reference factor.

These results also indicate some clear limitations of the LCS array in characterizing VOC sources. For example, the analysis of reference data revealed important VOC source profiles associated with the chemical tracer, cooking, and emissions from human occupants that are not rep-

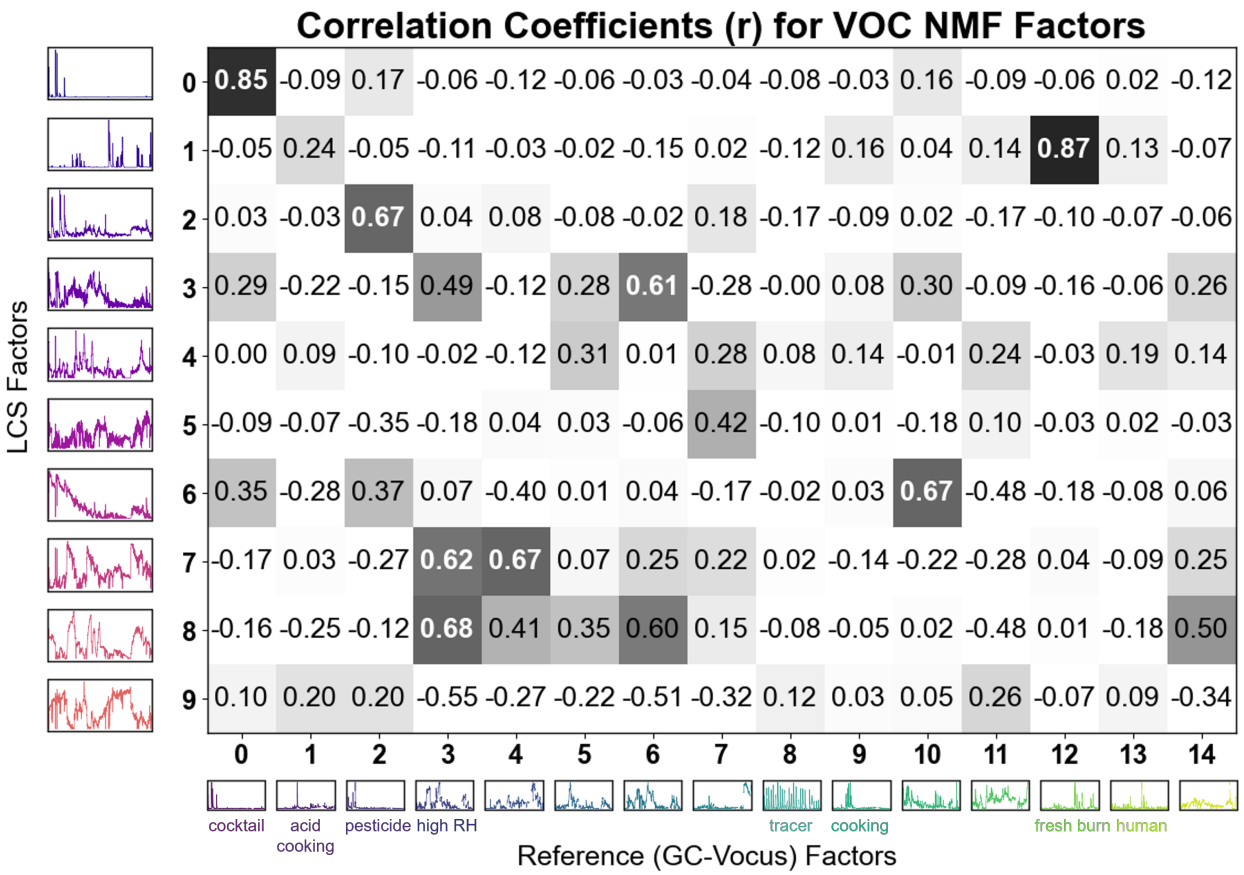

Figure 3.5 Correlations (Pearson r) between the output of the low-cost sensor factor analysis and the reference NMF factors, with shaded squares indicating higher positive correlation.

resented by the LCS-derived profiles. One possibility is that these reference profiles are caused by limitations of the NMF method, as factor analysis techniques sometimes "split" or "mix" the information from real factors depending on input parameters [101]. In Chapter 4, we show that information loss about these sources via NMF artifacts such as "splitting" or "mixing" is likely not the case for our low-cost signals. This implies that the lack of information about these sources is due to measurement issues: either LCS are not very sensitive to the compounds emitted from these sources, or these sources have concentrations that are very low relative to other sources. For example, the axes of Figure 3.1 show that all of these source profiles are at least an order of magnitude lower than that of the "chemical cocktail," which likely explains why our LCS analysis can't capture human emissions or cooking. On the other hand, the inability of our sensors to capture the "acid cooking" factor or tracer emission profile likely indicate a lack of sensitivity to the compounds emitted from these sources.

Finally, we should consider the limitations of the NMF method in capturing the behavior of all indoor VOCs. NMF produces factors based on the underlying co-variability of VOC time series signals. The variability of higher-volatility compounds is driven mostly by pollution sources, ventilation, and chemical reactions, while lower-volatility compounds, which tend to adsorb to indoor surfaces, have variability that is mostly dependent on changes to indoor environmental parameters [102].

When an indoor environment is very well-controlled, certain compounds could constitute a major source of indoor VOC pollution via passive emission from indoor surfaces or materials, but exhibit very little variability. In these cases, NMF and similar factor analysis techniques would likely fail to identify the importance of these longer-lived compounds to the indoor environment.

3.4 Conclusions and Future Work

By performing a NMF decomposition of low-cost VOC sensor signals and a correlation analysis of these results with reference-derived pollution source profiles, we have demonstrated that uncalibrated LCS data can yield interpretable insights into the sources and composition of indoor air pollution. Our LCS measurements provide meaningful information about chemical cocktail, fresh burning, and pesticide emission sources caused by human activities during the CASA indoor air campaign. The LCS also capture the behavior of RH- and temperature-driven changes in emissions of VOCs from indoor surfaces, as well as an unidentified VOC emission profile.

Our analysis also shows some clear limitations of using LCS in indoor air applications, as LCS do not give the complete picture of indoor VOC composition. For example, the analysis of reference data showed important VOC source profiles associated with cooking and emissions from human occupants that are not seen at all in the LCS-derived profiles. This is most likely due to either relatively low LCS sensitivities to the compounds emitted from these sources, or to the overall lower concentrations of these sources as reflected in the cps intensities of Figure 3.1. Low correlations with reference profiles could also theoretically be caused by limitations of the NMF method, whose outputs depend on the specific inputs and parameters utilized. In Chapter 4, we show that there are no combinations of sensors or NMF input parameters that yield meaningful information about these source profiles, ruling out this possibility.

Ultimately, we conclude that this low-cost approach will never be as quantitative as PTR-MS but seems feasible as a method to make relatively inexpensive, spatially-distributed measurements of indoor air. Overall, this work shows that, even without accounting for sensor interferences or achieving a comprehensive calibration, low-cost VOC sensors can be used for understanding sources of indoor pollution and providing insights into the underlying chemistry. This work is important as a proof-of-concept, and a similar approach could be used to characterize pollution sources with elevated VOC concentrations, such as outdoor wildfire or oil-and-gas emissions. We also expect that future work will utilize this approach to make spatially-distributed low-cost VOC measurements of indoor air. In Chapter 4, we optimize the sensor array to maximize NMF output and minimize the number of sensors, helping to reduce material cost and simplify data analysis for future spatially-distributed applications.

3.5 Acknowledgements

This work was made possible by funding from the Sloan Foundation, the funding source for both the development of our low-cost VOC sensor array and the CASA experiment. We thank the organizers of CASA – Nina Vance, Delphine Farmer, Dustin Poppendieck, and all the other participants and NIST staff who helped to coordinate the experiments. We also appreciate the hard work of the Abbatt group at UToronto in collecting and processing the VOC reference data; in particular, we

would like to thank Han Huynh and Jenna Ditto for their input on our reference data analysis. We also extend thanks to Amy Hrdina, who was responsible for making sure the low-cost VOC instruments remained up and running for the last few weeks of CASA measurements.

Appendix

3.A Appendix

3.A.1 Imputation Cross Validation Results for Reference VOC Data

For this dataset, we chose an NMF model that utilizes a coordinate descent solver to minimize Euclidean distance (Frobenius norm) between the original values and the factorized product W×H. We ran the CV2K imputation cross validation method 100 times at each possible rank k, for values of k between 1 and 17, on our filtered and scaled reference data. CV2K includes a parsimony condition to choose a rank when the cross-validation error has plateaued. In essence, we first find the rank with the lowest median, then perform a "rollback" by computing and comparing Wilcoxon rank-sum test[1] statistics for lower values of k until there is no longer an improvement in the test statistic [94]. The results from this imputation cross-validation are shown in figure 3.A.1, where the shaded area represents one median absolute deviation, and the dotted line indicates the calculated rank based on a parsimony condition. These results indicate that 15 is the best rank after the rollback process, but manual inspection of NMF solutions reveals that values of $k > 11$ also produced interpretable results.

3.A.2 Quantitative Identification of Reference Factors

Using the CASA experiment logs, we classified documented events in the house into one of 29 categories. These data were then transformed into binary time series, with 1s representing occurrence of an event and 0s representing no event (except in the case of the "People" time series, which was simply a scaled series of the number of house occupants at any given time). A quick summary of the different activity tags is given in Table 3.A.1.

We then used a peak finding algorithm from the scipy signal processing library to identify the peaks in each reference NMF factor. This series of peaks was then multiplied with the binary time series matrix, and the sums for each event were tabulated. Figure 3.A.2 shows an example of this process for Factor 0, which we can clearly distinguish as a factor representing the "chemical cocktail" pollution source.

In figure 3.A.3, we show the bar plots of products between NMF factor peaks and binary activity series for all 15 factors that were used to make assignments of pollution sources to factors. We know that several of these series correlate better with environmental variables (RH, T, dewpoint)

[1] The Wilcoxon rank-sum statistic tests whether two populations have the same median, with the null hypothesis being that the two populations have the same distribution and the same median [103]. This can also be thought of as a non-parametric t-test.

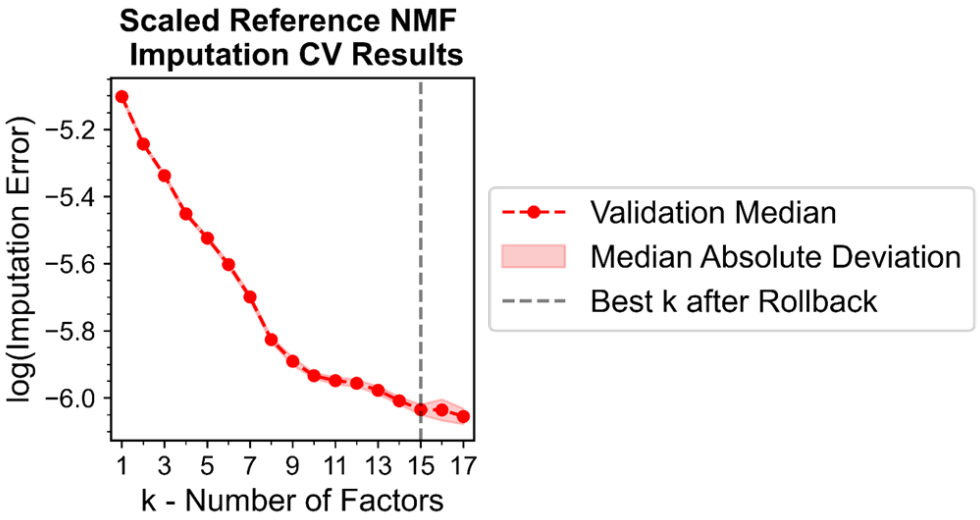

Figure 3.A.1 Results from CV2K imputation cross-validation on the scaled reference dataset. The shaded area represents 1 median absolute deviation, and the dotted line indicates the calculated rank based on a parsimony condition.

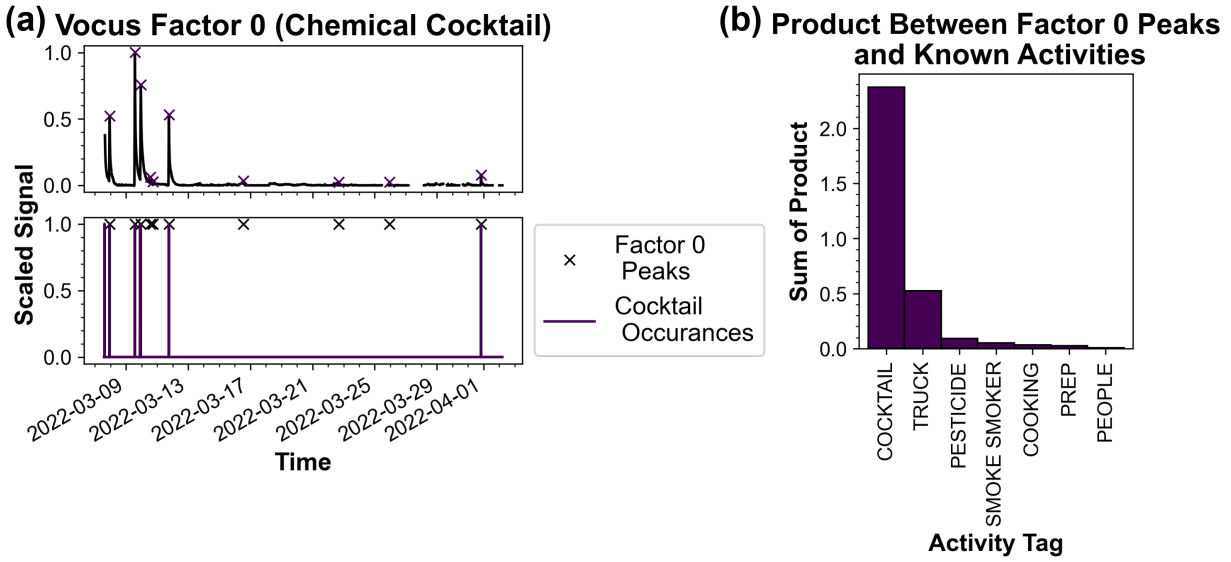

Figure 3.A.2 An example of the quantitative approach to associating reference factors with known activities and pollution sources. Peaks for Vocus factor 0 are identified and plotted in panel (a). The products of these peaks and the binary activity time series are then calculated and summed in panel (b). This NMF factor clearly represents the "chemical cocktail" source.

than they do with events in the house (see Appendix section 3.A.5). However, factors 7 and 10 have identifiable peaks that do not correlate well with any interpretable subset of the known activity series.

Table 3.A.1 Summary of activity tags for CASA events.

Name	Description	Name	Description
ACID	CO_2 injection during acid/base experiments.	PEOPLE	Time series of approximate number of humans in the house.
AIRFRY	Operation of air fryer to cook food.	PESTICIDE	Pesticide addition.
BASE	NH3 injection during acid/base experiments	POPCORN	Preparation of microwave popcorn.
COCKTAIL	"Chemical cocktail" injection.	PREP	Human occupancy in the house (maintaining instruments).
COOKING	Pan-frying any food (excluding vinegar).	REPAIR	Intermittent repair of a heat valve in the test house basement by handymen.
COOKING ACID	Pan-frying balsamic vinegar.	SMOKE BAG	Introduction of fresh smoke from the Teflon chamber.
COOKING PREP	Preparation of food for cooking by participants in the kitchen.	SMOKE BAG AGED	Introduction of aged smoke from the Teflon chamber.
DUSTING	Surface cleaning by dusting.	SMOKE PREP	Preparation of smoke experiments (filling smoke chamber).
HIGH RH	Maintenance of high RH in the absence of other activities.	SMOKE SMOKER	Introduction of fresh smoke from a cocktail smoker.
LAUNDRY	Automated laundry cycles, sometimes run with detergent.	TEMPRAMP	Periods of high temperature in the absence of other activities.
MOP	Mopping floors with cleaning solution.	TRACER	Nightly injection of SF_6.
OH	Addition of OH using OH generator.	TRUCK	Presence of an idling vehicle outside the test house.
OZONE	Addition of O_3 using ozone generator.	VACUUM	Vacuuming of floors.

3.A.3 Qualitative Compositional Data for Reference Factors

To determine the relative magnitude of each reference factor in cps, we fit an MLR model of the reference factor profiles to the total summed ion signal. This MLR model was cross-validated using a repeated holdout technique [104] with 70% of the data used for training and 20% used for testing. We found that the model had an r^2 of 0.99 when fit and applied to the whole dataset, and an average cross-validation r^2 score of 0.76 across 100 repetitions, indicating that the MLR model reported is relatively predictive and not a result of over-fitting.

Our NMF solution can give us information about the approximate chemical composition of each resolved factor. For each of the individual relevant VOCs in our dataset, we can find what percentage of its signal is due to a contribution from a factor, or pollution source. To simply this information,

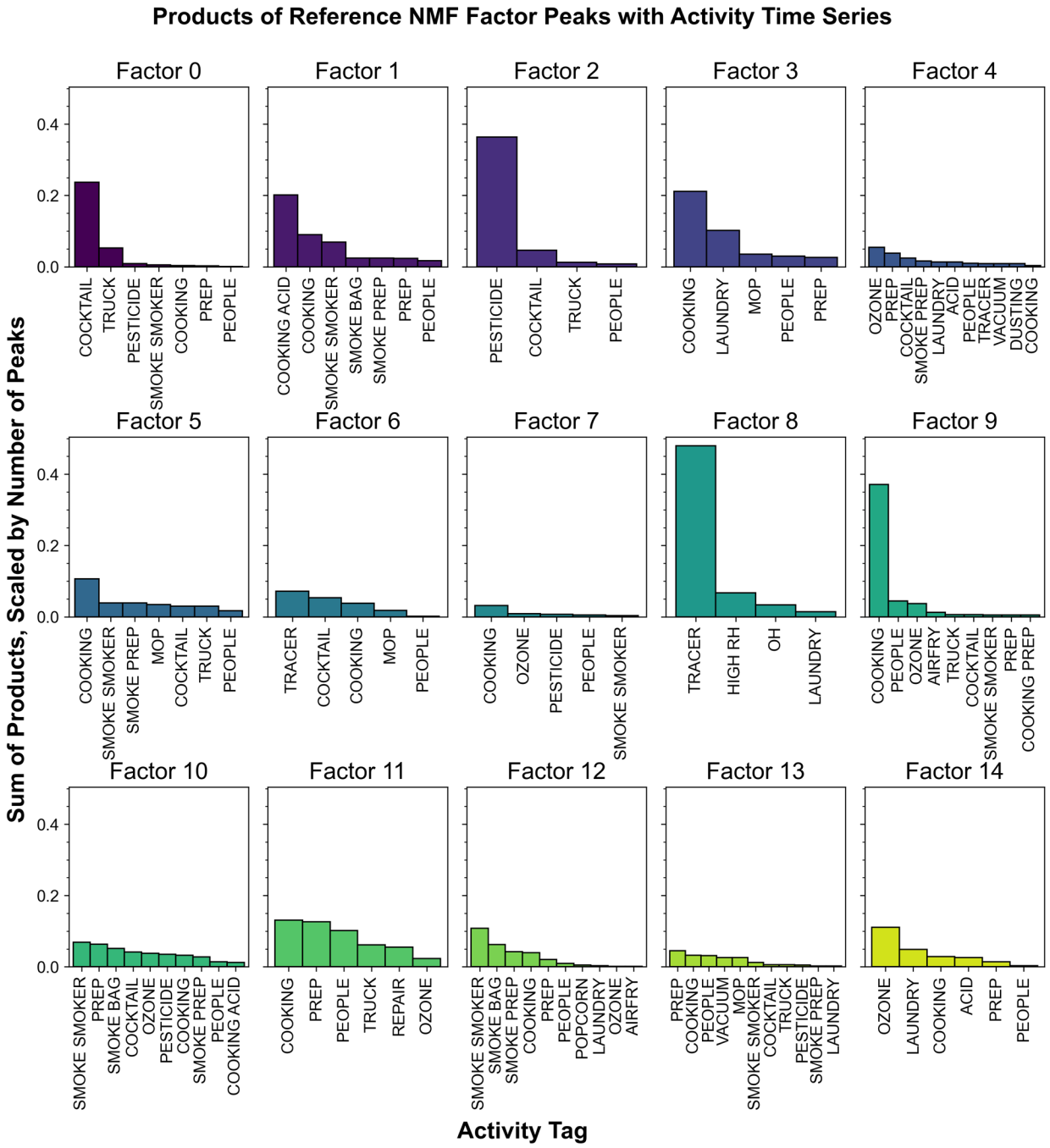

Figure 3.A.3 Products of the 15 reference NMF factors and the binary activity time series, summed and scaled.

we classified each of the VOCs by chemical composition, and then calculated the total amount of lumped signal explained by each factor. This gives us a rough "fingerprint" for each pollution source, as shown in Figure 3.A.4, with the caveat that the magnitude of the bars does not give us truly compositional information: rather, for every compound class, the sum of bars across plots equals one. For example, 25% of the total C_xH_y signal can be explained by factor 0, while Factor 13 explains

all of the total siloxane signal and small fractions of other compound classes.

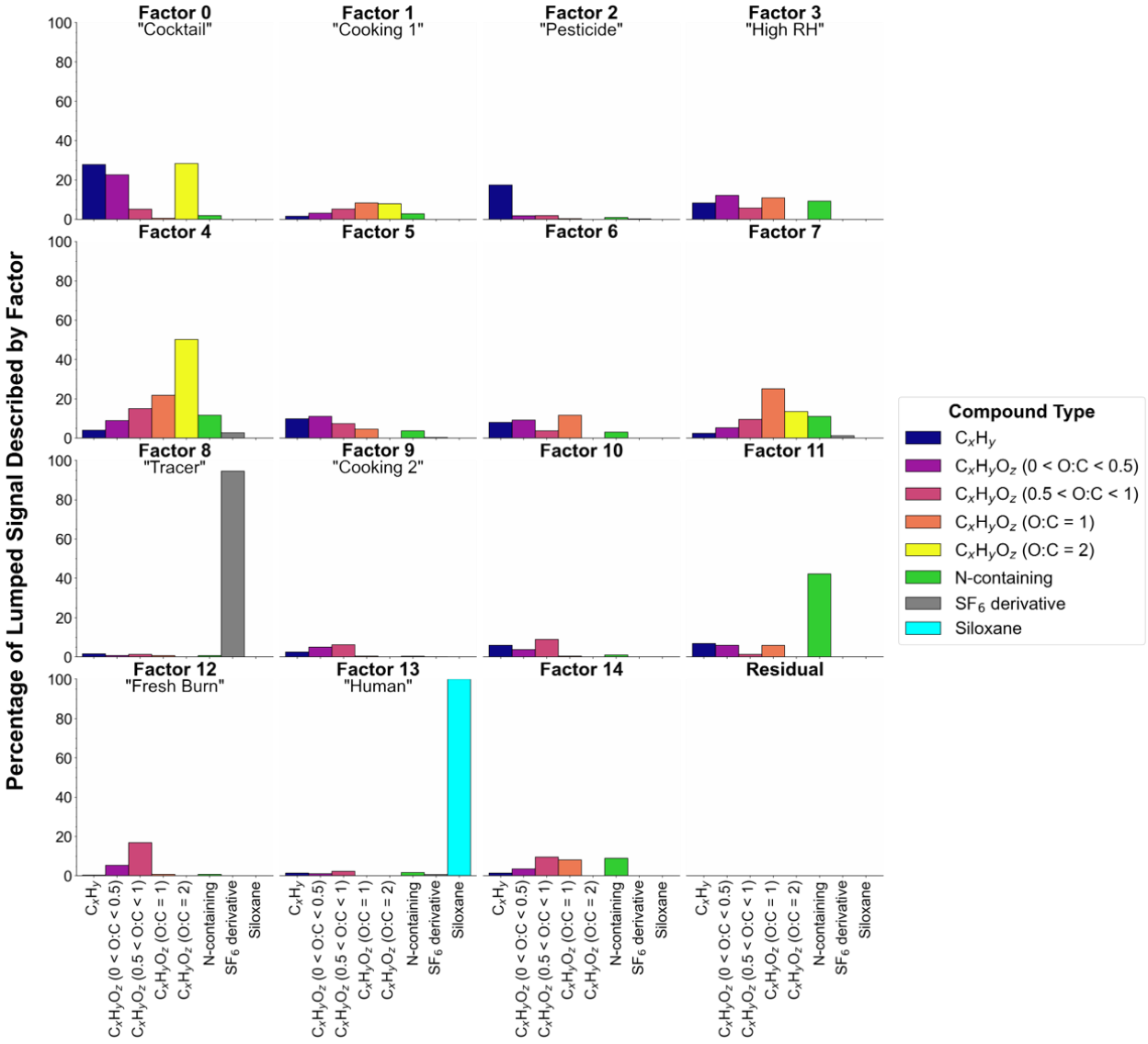

Figure 3.A.4 Fraction of each input species associated with a given reference NMF factor. For example, 25% of the total C_xH_y signal can be explained by factor 0, while Factor 13 explains all of the total siloxane signal and small fractions of other compound classes.

To explain how we arrived at the results shown in Figure 3.2 from the information in Figure 3.A.4, we use the following simplified example: suppose we perform a factor analysis of concentration time series for compounds A, B, and C, which have relative maximum concentrations of 4:2:1. From our factor analysis of the normalized time series for A, B, and C, we obtain 2 unitless factor time series, representing Source 1 and Source 2, that explain the behavior of A, B, and C without significant error. We also know or assume that Source 1 and Source 2 do not emit any other pollutants other than A, B, and C. Suppose that Source 1 is found to explain 50% of the scaled signal of compound A and 25% of the scaled signal of compound B. Because the measured concentration ratio of A to B is 2:1, this means that the ratio of A to B in contained in Source 1 is (50% × 2) : (25% ×

1) or 4:1. Similarly, Source 2 explains 50% of A, 75% of B, and 100% of C. This means that, based on the relative maximum concentrations of A, B, and C, Source 2 has an A:B:C ratio of (50%×4):(75%×2):(100%×1) or 4:3:2.

Thus in our dataset, if we assume that the sources (represented by NMF factors) emit only the VOCs we consider, and that ion signal in cps is directly proportional to concentration of VOC, then we can determine the fractional VOC composition of each factor based on the relative magnitudes of the original VOC ion signals.

3.A.4 Reference NMF Performed Without Scaling

In general, NMF models that utilize Euclidean distance metrics are not scale invariant, and either normalizing or standardizing the data to a common range is a standard preprocessing step [59]. However, we also chose to run NMF without scaling to compare results, as a related technique (Positive Matrix Factorization or PMF) that is commonly used with environmental data is not typically run with normalized data, despite also utilizing a Euclidean distance metric that is heavily affected by disparities in scale between input time series [60]. The results from a CV2K imputation cross-validation (with 100 runs at each possible rank) are shown in Figure 3.A.5. The NMF models for the data without scaling have much larger variance than those for the scaled data shown in Figure 3.A.1, though both cross-validation processes selected 15-factor solutions.

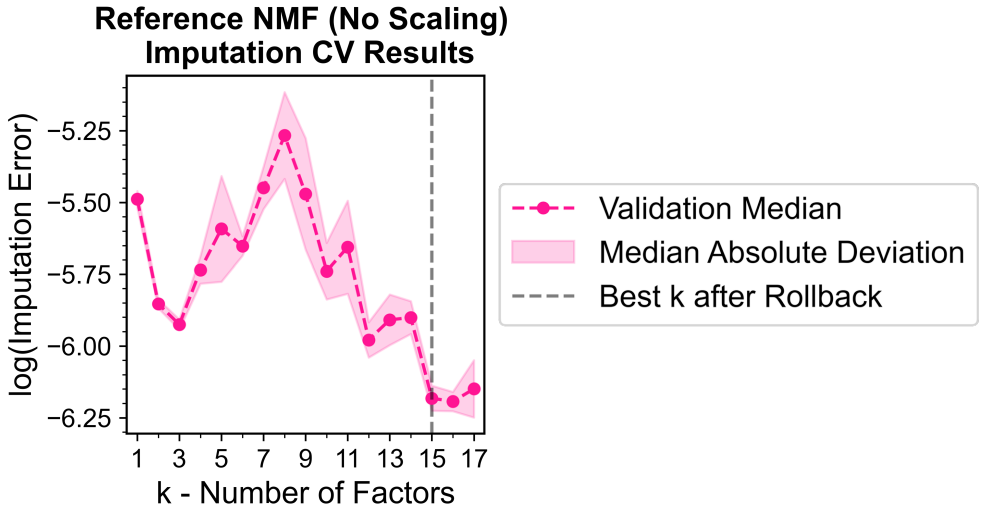

Figure 3.A.5 Results from CV2K imputation cross-validation on the reference dataset, without scaling. The shaded area represents 1 median absolute deviation, and the dotted line indicates the calculated rank based on a parsimony condition.

Figure 3.A.6 shows the 15-factor solution for the data without scaling. Many of these factors are physically interpretable, as they correspond to known activities in the house. Due to the relatively large concentrations of VOCs during the “chemical cocktail” experiments, this NMF solution has separated out 5 similar “chemical cocktail” emission trends as individual pollution sources. Moreover, it also prioritized single pollution events, such as mopping and microwave popcorn, that had an outsized effect on emissions of certain compounds (for example, the trend seen in factor 11 is almost entirely due to $C_3H_7O^+$, which is likely to be propanal and a known component of popcorn

aroma [105] and is the primary contributor to the trend seen in Factor 11). Ultimately, this solution seems far less representative of indoor VOC sources than the scaled solution reported in the main text.

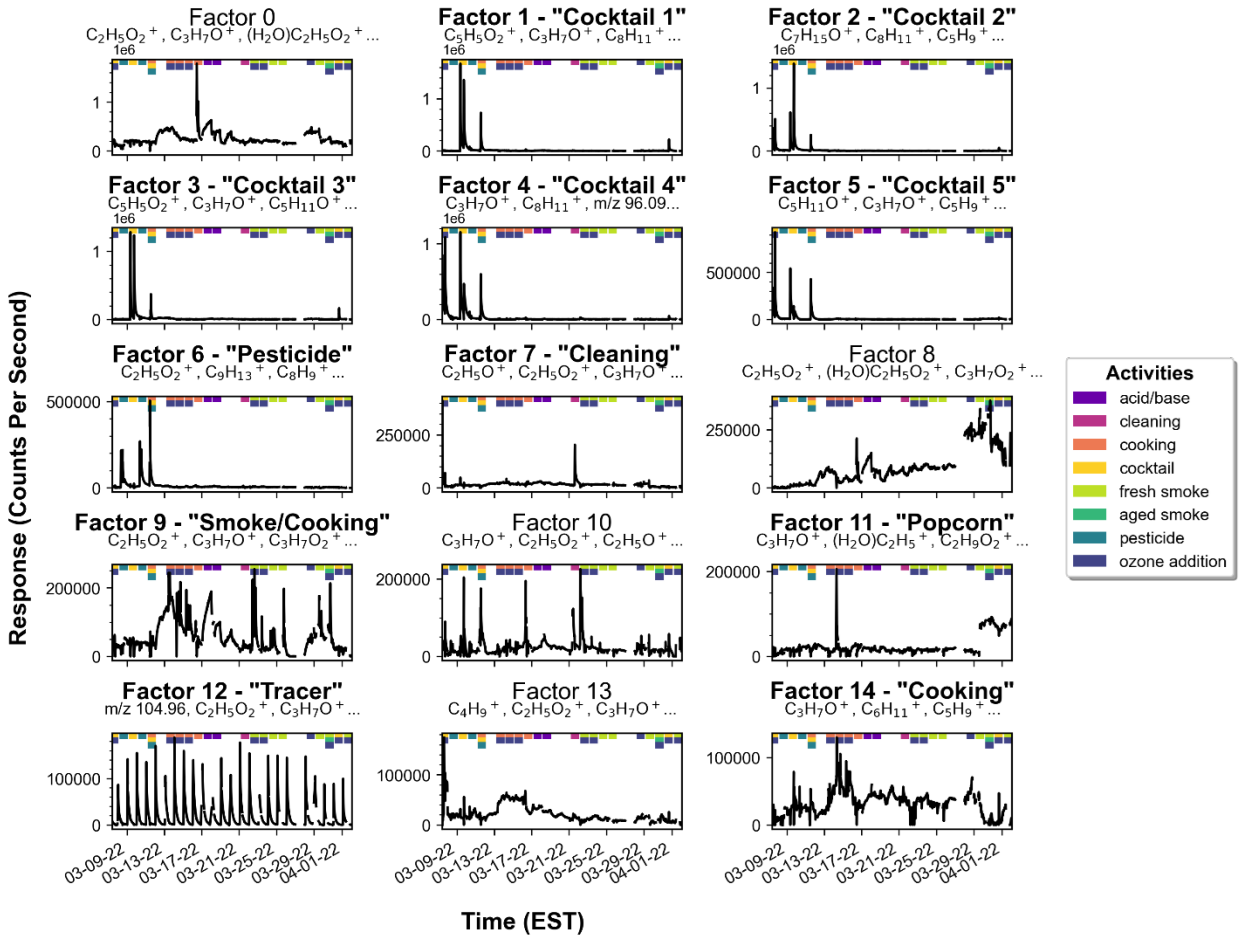

Figure 3.A.6 15-factor NMF solution for input reference VOC data without scaling. Colored dots indicate the primary activity for each day of experiments, as described in Table 3.1 The ions in the subtitles are the three ions that are most associated with the factor, in terms of how much of their signal is explained by that factor.

3.A.5 Correlations of Scaled Low-Cost Sensor Data

Figure 3.A.7 shows the normalized low-cost sensor signals used in our analysis. These plots include PM1, PM2.5, and PM10, which are Plantower PMS measurements that are not utilized in the main chapter of this text, but are included due to their relevance in Chapter 4. For descriptions of specific sensor measurements, see Table 3.2. Many of these sensor signals show moderate or strong correlations with other sensor signals, as seen in the cross-correlation matrix in figure 3.A.8. Figure 3.A.9 shows Pearson *r* correlations calculated between the NMF factors obtained from the reference Vocus dataset and the LCS measurements, including environmental parameters.

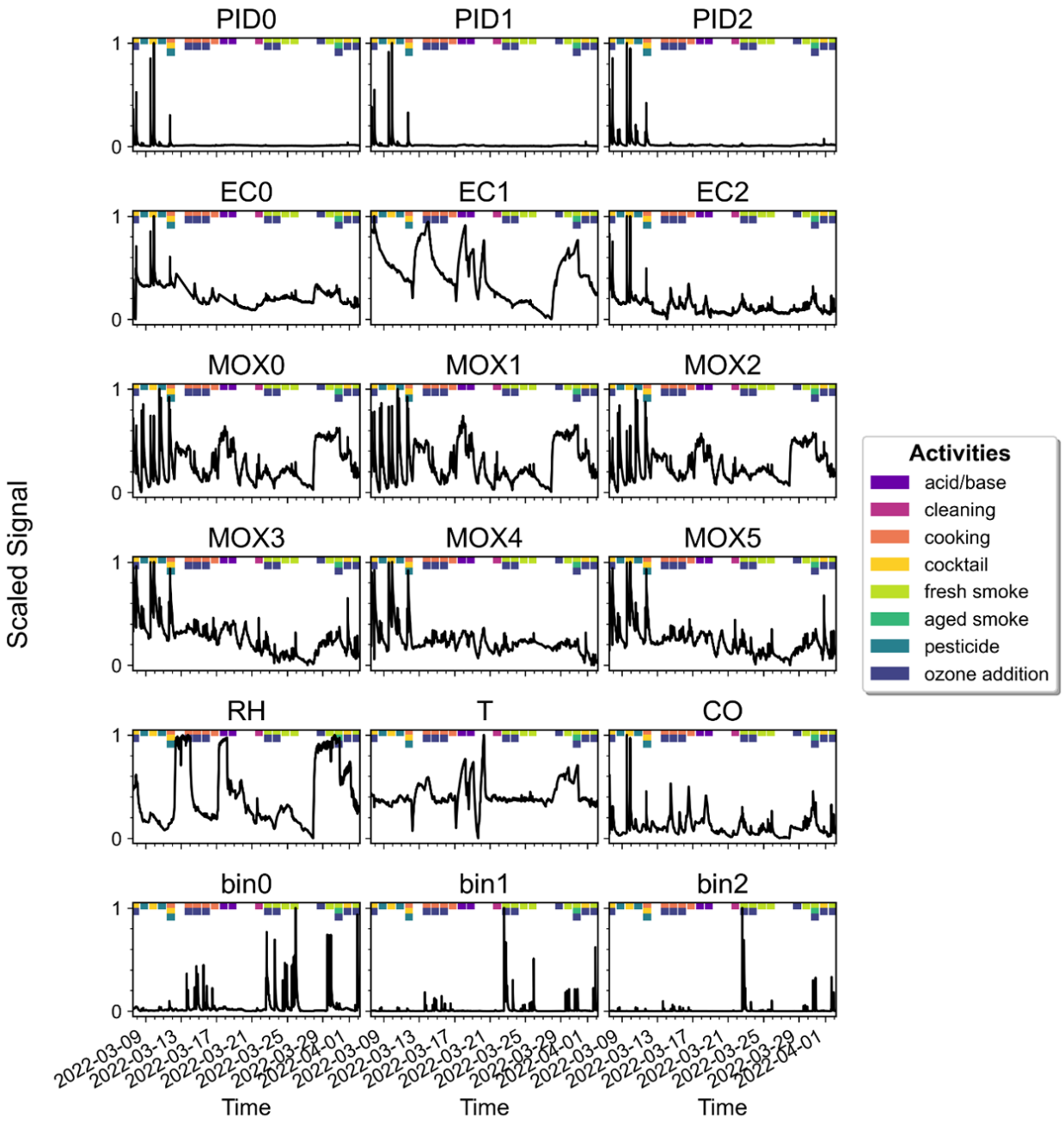

Figure 3.A.7 Scaled time series profiles for low-cost measurements.

3.A.6 Imputation Cross Validation Results for Low-Cost Sensor Dataset

We ran the CV2K imputation cross validation method, as described in sections 3.A.1, on our scaled low-cost sensor dataset for 100 runs at each possible rank k. The results of this imputation cross-validation are shown in Figure 3.A.10. The "rollback" process found an optimal rank of 6, but we found that 6, 8, and 10-factor solutions produced interpretable results. We chose the 10-factor solution, as this solution had a median imputation error close to the observed minimum and was also the lowest rank for which the LCS data were able to distinguish a "pesticide" factor.

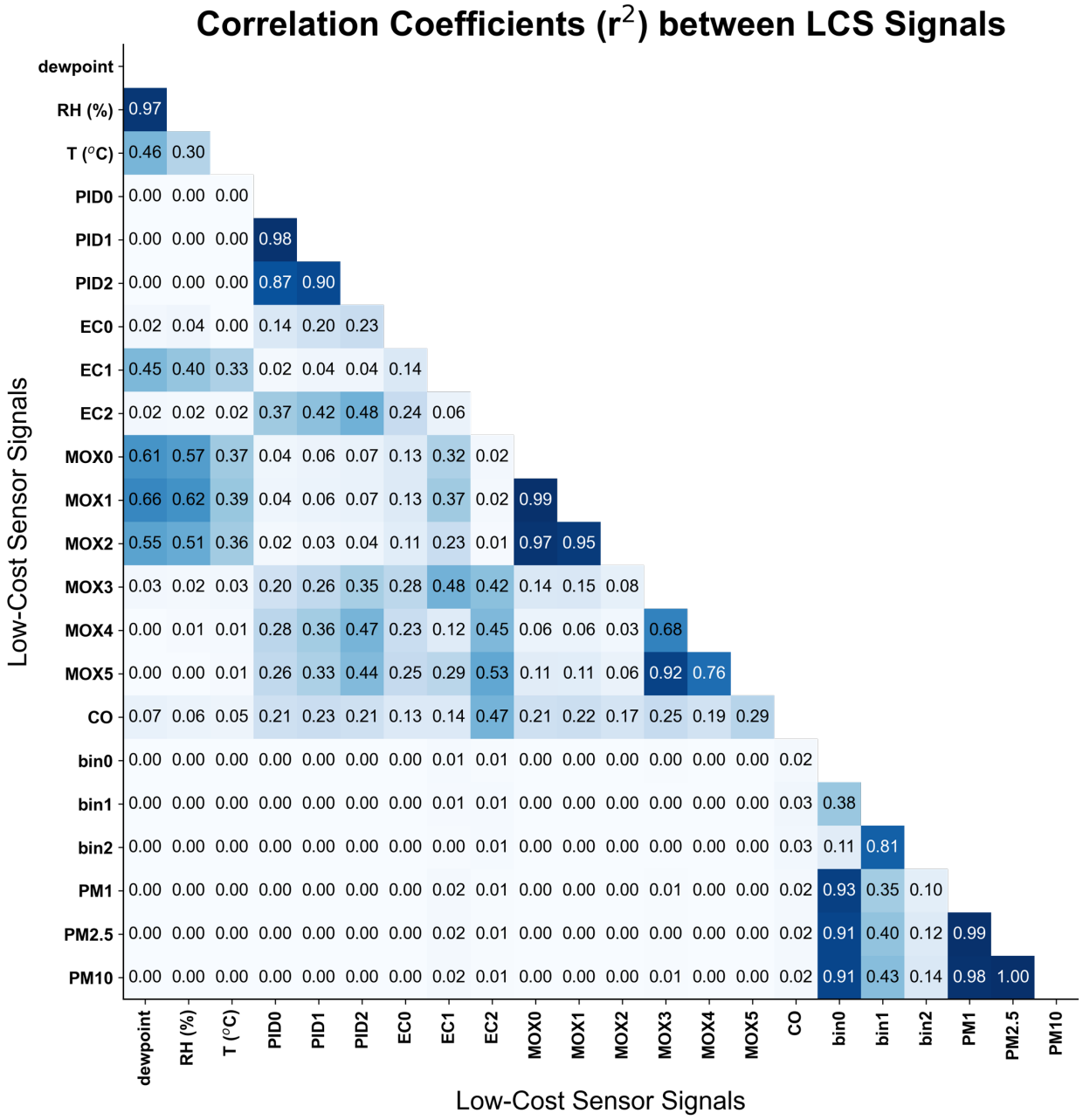

Figure 3.A.8 Cross-correlation matrix for all low-cost sensor signals, as described in Table 3.2. PM1, 2.5, and PM10 do not have descriptions in this chapter of the text, as they are not used in the described analysis, but are described in Table 4.2.1.

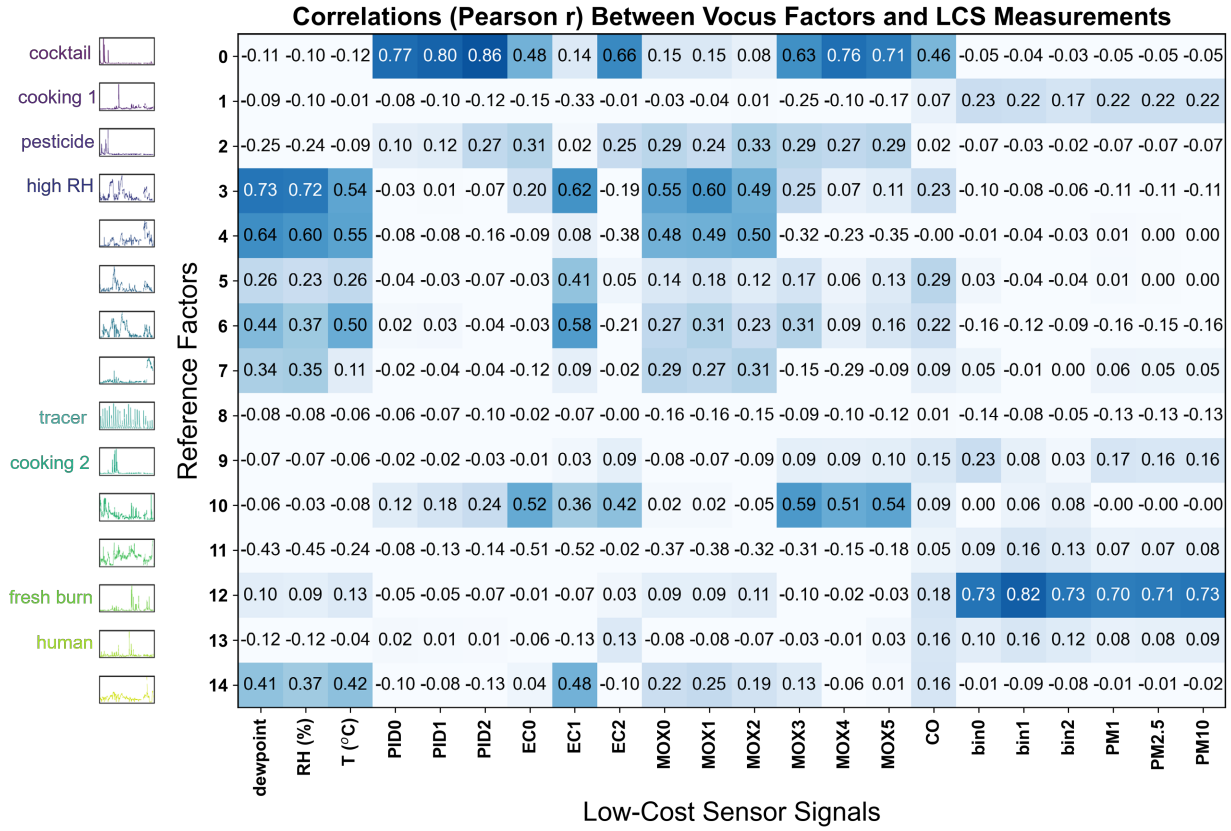

Figure 3.A.9 Correlations between the time series for reference NMF factors and measured low-cost sensor values, as described in Table 3.2.

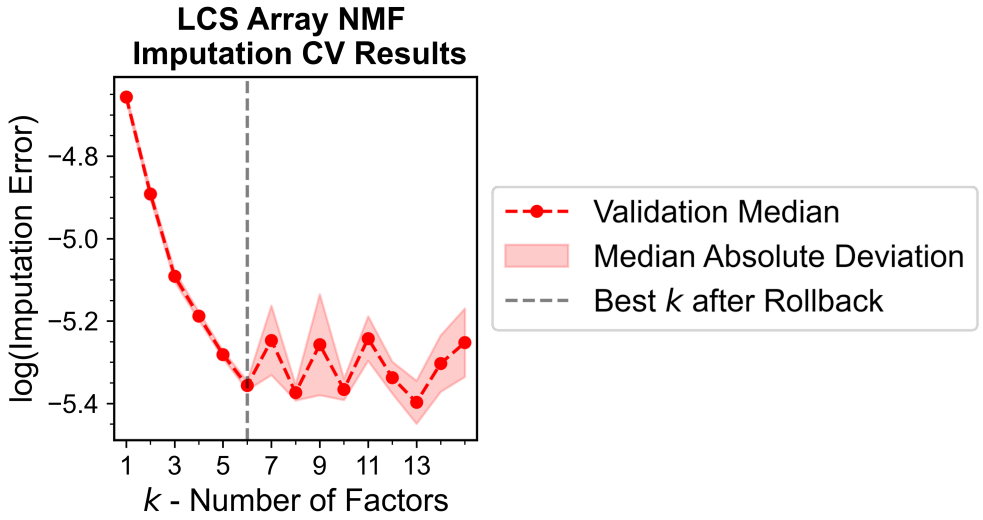

Figure 3.A.10 Results from CV2K imputation cross-validation on the LCS dataset. The shaded area represents 1 median absolute deviation, and the dotted line indicates the calculated rank based on a parsimony condition.

3.A.7 Activity Series Correlations for LCS-Derived Factors

In section 3.A.2, we described a quantitative method to identify the reference-derived NMF factors using binary series representing the activities in Table 3.A.1. We can use this same technique on the LCS-derived factors to associate them with known CASA activities. The results of this analysis are shown in Figure 3.A.11.

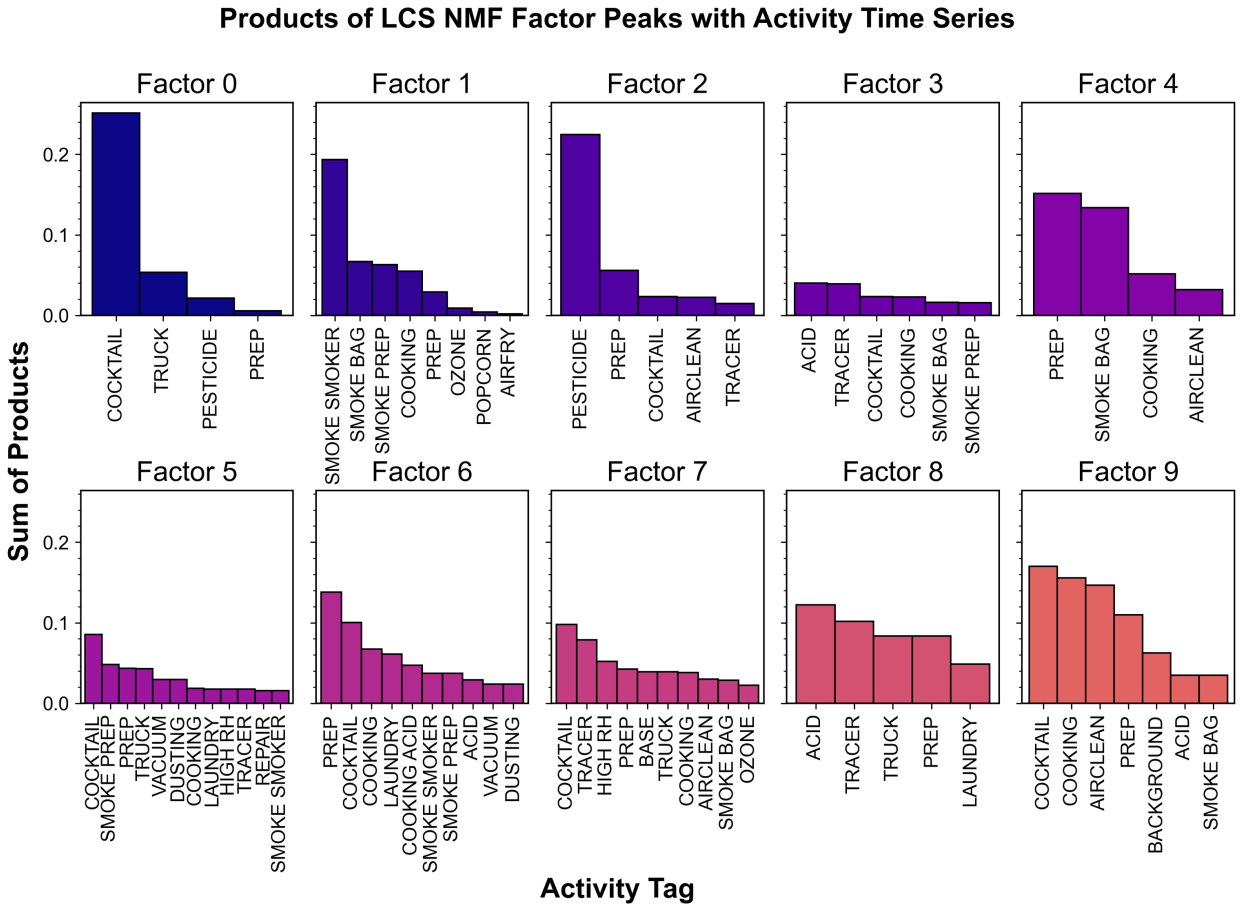

Figure 3.A.11 Quantitative identification of LCS-derived NMF factors, based on the summed and scaled products between binary activity series described in Table 3.A.1 and identified peaks for each factor.

This method independently yields accurate identifications of LCS-derived factors representing the source profiles for "chemical cocktail" emissions (LCS Factor 0), fresh burning (LCS Factor 1), and pesticide usage (LCS Factor 2). These results are consistent with the factor identifications found using the correlation analysis described in the main text and shown in Figure 3.5. In other words, in the absence of reference-grade VOC data, the activity series of known VOC emission events can be used to interpret the results of a LCS-derived NMF analysis.

3.A.8 LCS and Chemical "Fingerprints" for Correlated Factors

In Figure 3.A.12, we have juxtaposed the chemical factor compositions from Figures 3.2 and 3.4 for pairings of factors with r>0.6. This allows us to see some trends between compounds emitted

from sources and the sensor signals that are associated with them.

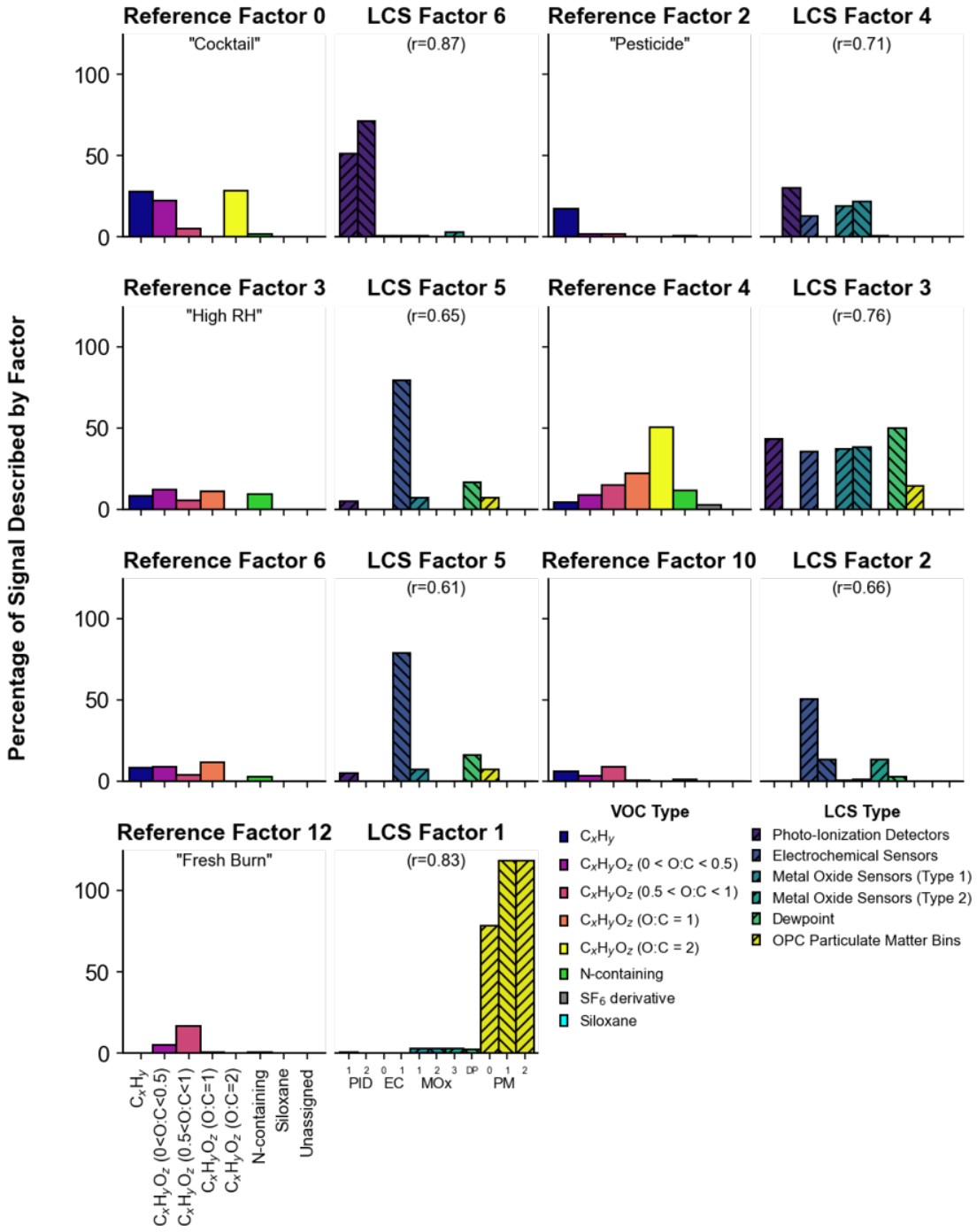

Figure 3.A.12 Juxtaposition of the reference factor chemical compositions with the fractions of LCS signal explained by corresponding LCS factors for any pairings of factors with $r > 0.6$.

Chapter 4

Assessing Low-Cost Multi-Pollutant Array Configurations for Measuring VOC Sources

4.1 Introduction

In Chapter 3, we showed that application of non-negative matrix factorization (NMF) to an array of low-cost measurements, including 12 low-cost VOC measurements, can provide meaningful information about indoor VOC sources. However, it seems likely that not all sensor signals are equally useful for this analysis, and it is important to know which low-cost measurements contribute the most information about VOC sources as this knowledge could lead to the future development of a VOC instrument, with a smaller array of targeted sensors, that is more suited for spatially-distributed applications. Solving this problem of sensor selection, or determining the smallest combination of sensors that can detect and identify the compounds of interest, reduces the material cost of the array and prevents redundancies and multi-collinearities in the input data that could interfere with data analysis [106]. To optimize sensor array configuration, Corcoran suggests a "structured approach" that involves the specification of a cost function, which should quantify the performance of a sensor sub-array in the application of interest, and a search algorithm, which should minimize this cost and output an optimal or near-optimal sensor configuration [107].

Sensor selection processes have commonly been being used for “electronic nose” applications, where a gas sensor array is employed to detect high concentrations of VOCs in odors or gas mixtures [36]. Some of these approaches are exhaustive, such as the one described in Chaudry et al.: to select a subset of 20 polymer sensors from an original group of over 200, the researchers developed sensitivity and selectivity metrics for sub-arrays, then used a leave-one-out approach to remove the single "worst" sensor (or the sensor whose inclusion led to the least improvement in metrics) until the target number of sensors was reached [108]. Other studies have taken unique approaches to defining a cost function for sensor arrays, such as when Zhang et al. developed a sensor selection method for six MOx sensors measuring 11 different VOCs at 100-400 ppm: they split the dataset into small recognition tasks between two gases and then determined the optimal sensor sub-array for each task, finding an optimized array of three sensors that could successfully distinguish between all pairs of gases [109]. Finally, some researchers have avoided an exhaustive combinatorial search through their choice of search algorithm, such as when Wei et al. sought to select a sub-array of six MOx sensors measuring pulses of 12 different combinations of CO and

methane with a random forest approach that found an optimized sub-array of two sensors. [110]. In all these cases, the dimensionalities of the measurement arrays were reduced significantly without much loss in the ability of the arrays to distinguish between and quantify the compounds of interest.

These past sensor selection studies have two clear limitations that are relevant to our work. The first is that these studies are focused on optimizing sensors with the same fundamental measurement principle, while our array contains three distinct measurement technologies and distinct sensor partial sensitivities achieved within each measurement type. Results from Chapters 2 and 3 suggest that this measurement multidimensionality is beneficial, and the process of sensor selection will allow us to quantitatively examine this assumption about the importance of multiple measurement types and diversity across individual sensors. The second limitation is that past sensor selection studies optimized sub-array performance based on laboratory data, rather than real-world VOC data. Similarly, we could attempt to optimize sensor configurations based on the the laboratory results from Chapter 2. However, as we discussed in Section 2.4, laboratory results do not necessarily translate well to real-world applications of these sensors. Instead, the NMF correlation analysis discussed in Chapter 3 provides a better framework for exploring relative sensor importance, as it produces quantitative metrics for the sensor array's ability to capture information about real-world sources of VOCs.

Here, we describe a brute-force, combinatorial approach, where the factor and correlation analyses described in Chapter 3 are repeated for an exhaustive number of sensor sub-arrays. A numeric score, calculated from an arbitrary but relevant cost function, can be determined for each sensor sub-array. This metric can be exploited to find optimal combinations of sensors, allowing us to assess the relative importance of individual sensors or groups of sensors in characterizing indoor VOC sources. These results demonstrate the robustness of our NMF results and could help to inform future applications of low-cost sensors for measuring environmental VOC sources.

4.2 Methods

4.2.1 Low-Cost and Reference Measurements

Here, we utilize the same measurements that are described in detail in section 3.2.1. Measurements were taken at the Chemical Assessment of Surface and Air (CASA) campaign, a collaborative indoor field experiment conducted at a model two-story residential home. Several specific perturbations to indoor air quality and chemistry were performed, as we previously summarized in Table 3.1. Low-cost measurements of VOCs and environmental parameters (relative humidity and temperature) were made by the low-cost VOC sensor array described in Chapter 2. Briefly, this is an array of 12 low-cost sensors, representing three fundamentally different sensing technologies: photo-ionization detection (PID, 3 sensors total), electrochemical (EC, 3 sensors) sensing, and metal oxide (MOx, 6 sensors) sensing. We were able to vary operational or physical parameters between sensors with the same sensing technology such that each of the 12 sensors has its own distinct set of sensitivities to various VOCs. This analysis also includes co-located low-cost measurements of CO and particulate matter (PM) made using a QuantAQ Modulair air quality monitor. The Modulair uses an Alphasense CO-B1 sensor to make measurements of CO and uses two different low-cost optical particle sensors (Alphasense OPC-N3 optical particle counter and Plantower PMS nephelometer) to make size-resolved particle measurements. For the following analysis, it is

useful to separate our low-cost sensor measurements into four categories: VOC, environmental ("Env"), CO, and particulate matter ("PM"). We have summarized the low-cost measurements that comprise each category in Table 4.2.1.

Table 4.2.1 Summary of LCS measurements used in configuration analysis.

Name	Type	Description	Name	Type	Description
PID0	VOC	ION Science MiniPID 2 10.0eV	RH	Env	Relative humidity (Sensiron SHT25)
PID1	VOC	ION Science MiniPID 2 10.6 eV	T	Env	Temperature (Sensiron SHT25)
PID2	VOC	ION Science MiniPID HS 10.6 eV	DP	Env	Dewpoint, calculated from RH/T
EC0	VOC	Alphasense VOC-B4 VOC sensor (no bias)	CO	CO	Alphasense CO-B4 sensor
EC1	VOC	Alphasense VOC-B4 sensor (positive bias)	bin0	PM	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.38-0.46 µm
EC2	VOC	Alphasense ETO-B4 sensor	bin1	PM	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.46-0.66 µm
MOX0	VOC	Figaro TGS2600 or "Type 1" MOx (5.0 V supply voltage)	bin2	PM	Alphasense OPC-N3: binned number concentration of particles with diameter between 0.66-1.0 µm
MOX1	VOC	Figaro TGS2600 (4.75 V supply voltage)	PM1	PM	PM_1 estimate from Plantower PMS nephelometer
MOX2	VOC	Figaro TGS2600 (5.25 V supply voltage)	PM2.5	PM	$PM_{2.5}$ estimate from Plantower PMS nephelometer
MOX3	VOC	Figaro TGS2602 or "Type 2" MOx (5.0 V supply voltage)	PM10	PM	PM_{10} estimate from Plantower PMS nephelometer
MOX4	VOC	Figaro TGS2602 (4.75 V supply voltage)			
MOX5	VOC	Figaro TGS2602 (5.25 V supply voltage)			

Our reference VOC measurements were made by a GC-Vocus PTR-ToF-MS, which is described in more detail in section 3.2.1. In section 3.3.1, we show that the reference VOC measurements can be meaningfully decomposed into 15 source profiles representing pollution emission trends during the CASA campaign. These 15 source profiles represent our "ground truth" for indoor VOC emissions during CASA, and the NMF results of our low-cost sensor array can be compared to these source profiles via a correlation analysis that is described in section 3.3.2.

4.2.2 Combinatorial Sensor Array Configuration Analysis

We considered many sub-arrays of sensor measurements, with the possible configurations detailed in Table 4.2.2. These configurations also include sensor sub-arrays where an entire category or many categories of sensors is omitted completely, including cases where VOC sensors and/or ancillary measurements (Env, CO, PM) were excluded. In theory, we could significantly reduce computation time by reducing the number of sub-arrays explored with a decision tree approach such as the one used in Wei et al [110]. However, despite its inefficiency, this brute-force approach of the entire parameter space allows us to arrive at more confident conclusions about optimal sensor array configurations, while also enabling examination of key features of the sensitivity of the array

such as the role of each VOC sensor type, the importance of varying sensor operational parameters, and the role of non-VOC ancillary measurements.

Table 4.2.2 Summary of all sensor array configurations considered.

Type	Combinations Considered	Total Number of Cases
VOC	No VOC sensors Any combination of n VOC sensors with $1 \leq n \leq 12$	4096
Env	No environmental parameters DP only RH and T only RH, T, and DP	4
CO	No CO CO	2
PM	No PM bin0 only OPC Measurements (bin0, bin1, bin2) only Nephelometer measurements (PM1, PM2.5, PM10) only bin0 + PM1 All OPC bins and nephelometer values	6
Total		196608

For each possible sub-array, the process described in section 3.3.2 was performed. The scaled responses from the sub-array from March 7 to April 2, 2022, were averaged to 10 minutes and represented as an input matrix of size $n \times m$, where n is the number of low-cost measurements in the sub-array and $m = 3691$ is the number of observations. Using the scipy implementation of NMF, run with a Euclidean beta-distance metric (Frobenius norm), we calculated the NMF results for the sub-array at every rank k, with $2 \leq k < \mathrm{n}$. Then, we performed a correlation analysis, calculating the Pearson r coefficient between every LCS-derived NMF factor and the 15 reference VOC source profiles.

We then assigned each correlation matrix a score that roughly represents the similarity between low-cost factors and the reference source profiles. In our primary scoring scheme, we consider only pairings of LCS factors and reference VOC source profiles that have a Pearson $r \geq 0.6$ (considering only the pairing with maximum correlation if there are multiple pairings with the same LCS factor or reference source profile), and simply take the sum of these correlations. To account for the diminishing returns with increasing NMF rank, we can calculate a new, rank-scaled score representing the improvement of a higher-factor solution over the 2-factor solution, divided by the number of factors k:

$$\mathrm{Score}_{\mathrm{rank-scaled}} = \frac{\mathrm{Score} - \mathrm{Score}_{\mathrm{k=2}}}{\mathrm{k}}. \tag{4.1}$$

In Section 4.3.2, we examine other metrics that include raising the threshold for correlations to $r \geq 0.7$; maximizing correlations with source profiles for "chemical cocktail," pesticide, and fresh burning emissions; and summing only correlations of $r \geq 0.6$ with "spiky" (entropy > 0.5) reference source profiles.

4.3 Results and Discussion

4.3.1 Optimized Sensor Array Configurations

Figure 4.3.1a shows the maximum score achieved amongst all possible sensor sub-arrays, calculated at every possible NMF rank. The sensor configurations for each rank that result in these maximum scores are given in Table 4.3.1. Figure 4.3.1b shows the improvement of higher-factor solutions over the 2-factor solution, divided by the number of factors, as given by Equation 4.1. In appendix section 4.A.1, we justify the use of this rank-scaled metric by examining a higher-factor solution and showing how it is not more informative than the optimal sensor configurations calculated for lower NMF ranks. We see that even though the score in Figure 4.3.1a reaches its maximum at a rank of $k = 11$, Figure 4.3.1b shows a maximum at $k = 6$. This suggests that, past 6 factors, the relative amount of information added by each additional factor is significantly diminished.

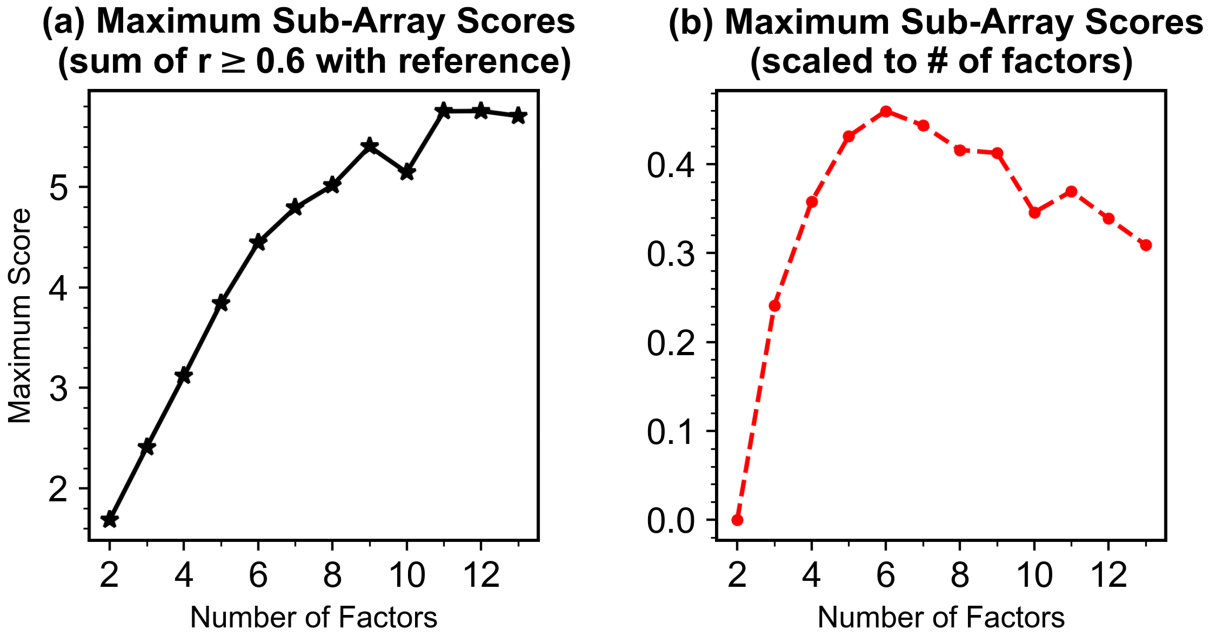

Figure 4.3.1 (a) Maximum score (sum of correlations above 0.6) achieved by any sub-array of sensors for ranks between 2-13 (b) The scores in panel a, modified to account for the diminishing returns of higher-factor solutions as in Eq. 4.1.

The sensor array configurations in Table 4.3.1 allow us to make a few observations about the relative importance of sensors. Every sensor sub-array contains at least 1 PID sensor, cementing the importance of this technology in the array. In addition, every sub-array contains PM data, and most combinations utilize size-resolved data; because VOCs and PM are co-emitted from the "fresh burning" sources, the inclusion of PM data helps the NMF analysis to extract information on this source profile. Environmental parameters also appear in nearly every optimal combination, likely due to the moderate dependence of indoor surface emissions on environmental parameters that we discussed in Section 3.3.1. MOx sensors are also important in sub-arrays optimized for greater than 3 factors, and the majority of these configurations include a mixture of MOx signals

of both Type 1 and Type 2 sensors. On the other hand, EC sensors, including the low-cost CO measurement, seem less critical for this analysis – while they are included in some optimal sensor configurations, they are not present in k = 6 or 7 sub-arrays that, based on Figure 4.3.1b, should be the most informative.

In summary, Table 4.3.1 shows that each "optimal" sensor configuration includes multiple VOC sensing technologies except for those optimized for very low-rank solutions (k = 2,3). These optimal configurations also suggest that inclusion of a low-cost CO measurement is not necessary, despite VOC EC sensors' cross-sensitivity to CO. In addition, inclusion of dewpoint or the combination of RH and T helps to maximize correlations with the reference dataset. Finally, size-resolved PM data is present in most of these configurations, but inclusion of data from both types of optical particle sensor does not appear to be necessary.

Table 4.3.1 Optimal sensor configurations for each possible rank.

Low-Cost Sensor Measurements		Number of Factors: k =	2	3	4	5	6	7	8	9	10	11	12	13
	VOC	PID0			✓		✓	✓		✓	✓			✓
		PID1	✓	✓	✓			✓	✓	✓	✓	✓		✓
		PID2	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓
		EC0							✓	✓		✓	✓	✓
		EC1				✓			✓		✓		✓	✓
		EC2									✓	✓	✓	✓
		MOx0			✓		✓		✓	✓		✓	✓	
		MOx1					✓	✓	✓			✓	✓	✓
		MOx2			✓	✓	✓		✓	✓	✓	✓		✓
		MOx3					✓	✓		✓	✓	✓	✓	✓
		MOx4						✓			✓		✓	✓
		MOx5					✓	✓		✓	✓	✓	✓	
		CO						✓		✓	✓	✓	✓	✓
	Env	DP		✓	✓	✓	✓			✓		✓		✓
		RH						✓	✓		✓		✓	
		T						✓	✓		✓		✓	
	PM	bin0	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
		bin1	✓	✓	✓	✓	✓	✓						
		bin2	✓	✓	✓	✓	✓	✓						
		PM1							✓		✓		✓	✓
		PM2.5									✓			
		PM10									✓			

This approach allows us to find sensor configurations that, when subjected to NMF, provide relatively high amounts of information. In Figure 4.3.2, we show the six-factor NMF solution for the highest-scoring sub-array for k=6, as shown in Table 4.3.1. This sub-array includes only 6 VOC sensors, and omits EC sensor signals while including ancillary measurements of DP and size-resolved OPC data. We can see that all six resolved factors line up well with VOC source profiles: LCS factors 0 and 2 correspond to two surface emission factors, while factors 1,4, and 5 are highly correlated with emissions from fresh burning, pesticide, and "chemical cocktail" respectively (LCS factor 3 correlates with unidentified VOC source profile 10, which we see from Figures 3.2 and

3.A.3 is primarily made up of non-oxidized C_xH_y compounds and may be a secondary burning factor). Figure 4.3.3 shows the percentage of each input low-cost sensor signal associated with a particular NMF factor. Much like the solution we saw in Section 3.3.2, we see that every factor is associated with at least a small fraction of VOC sensor response. We also see that the sub-array has similar factor compositions to the ones of the larger array shown in Figure 3.4, and notice that there are meaningful differences in the proportion of sensor responses explained by each factor. For example, Factor 5, which is correlated with the "chemical cocktail" source profile, is primarily associated with PID sensors while Factor 4, which is correlated with the pesticide source profile, is associated with PID and MOx Type 1 signals.

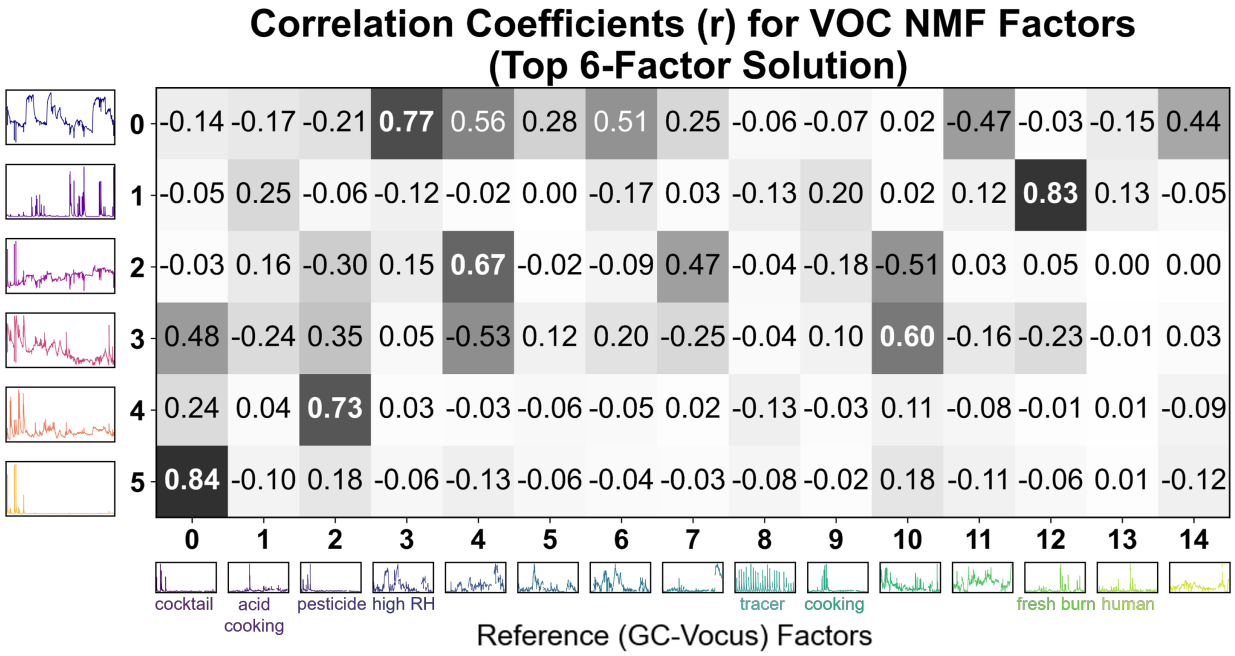

Figure 4.3.2 Correlations (Pearson *r*) between LCS-derived NMF factors and reference VOC source profiles, for the *k*=6 "optimal" sensor configuration, with darker squares indicating higher positive correlation.

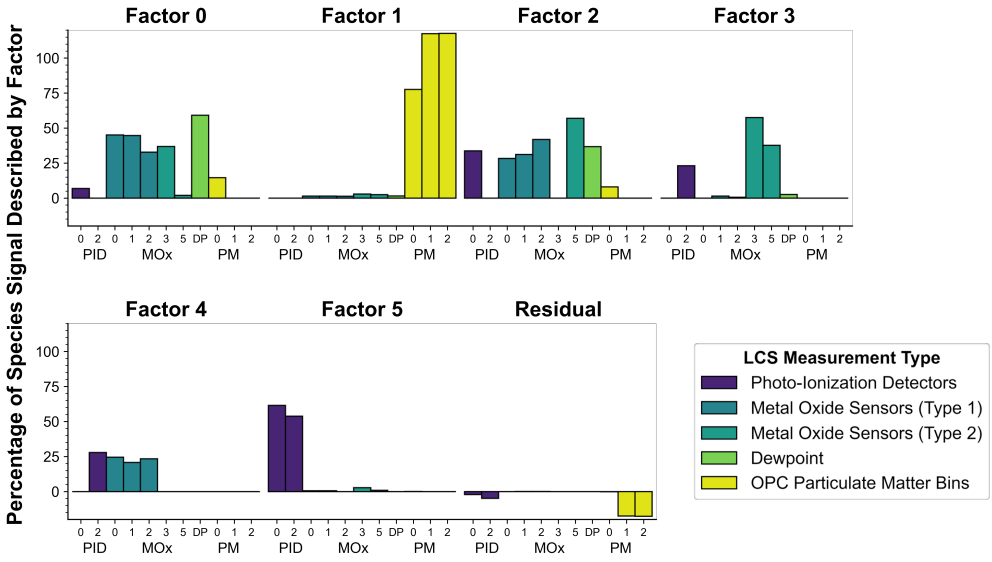

Figure 4.3.3 Fraction of each input low-cost sensor signal associated with a given LCS-derived factor, for the *k*=6 "optimal" sensor configuration.

To assess the relative importance of individual measurements or groups of measurements on the factorization results, we can plot these scores for a variety of scenarios. In Figure 4.3.4, we show the maximum rank-scaled score for sub-arrays with imposed restrictions, such as sub-arrays that are only allowed to contain VOC sensors or sub-arrays that are forced to omit environmental parameters (we show comparisons for the original score, or sum of correlations greater than 0.6, in appendix figure 4.A.3). In Figure 4.3.4a, we explore sub-arrays that omit various sensor measurements entirely. In the most extreme case, we withhold low-cost VOC measurements from possible sub-arrays. These constrained sub-arrays have low maximal scores, suggesting that low-cost VOC sensors are key for inferring these indoor pollution sources and ancillary measurements alone cannot accomplish this task. We see that omission of either Env or PM data results in only minor decreases in sub-array performance, but omitting both and using only low-cost VOC measurements does cause some information loss. This is because the inclusion of PM data helps to extract information about the fresh burning source profile, which is moderately correlated with PM data, while inclusion of environmental data helps to extract information about surface emissions, which are moderately correlated with environmental parameters (see Figure 3.A.9).

In panel Figure 4.3.4b, we show the maximal scores for sensor sub-arrays that are allowed to use PM and Env data but are constrained to one VOC measurement type. All of these curves are sub-optimal compared to the maximal scores of sub-arrays with no constraints, and this result confirms a key hypothesis underlying this work: that multiple VOC sensor technologies improve the ability of a sensor array to give useful information about VOC sources and composition. This hypothesis gains further support from Figure 4.3.4c, where we explore the performance of sub-arrays that must include three specific sensors representing each of the three sensing technologies. As a point of comparison, we also include the maximal performance of sub-arrays that are constrained to each of these specific VOC sensors. None of the sub-arrays containing a single VOC sensor have very high performance, but the sub-arrays containing all three sensors fare dramatically better. This suggests that including diverse measurement technologies helps to ensure that the sensor array captures the various indoor VOC sources, but we also note that most of the optimized sensor arrays in Table 4.3.1 have some degree of redundancy caused by including multiple sensors of each measurement type.

Finally, Figure 4.3.4b shows that the performance of sub-arrays limited to MOx sensors appears to be significantly higher than those of PID and EC-limited sub-arrays. To understand this observation, we plot the performance curves for sub-arrays that are limited to only MOx Type 1 and MOx Type 2, respectively, in panel (d). Both curves show significantly worse performance than the curve for sub-arrays that can utilize both types of MOx sensors. This result supports the findings of Collier-Oxandale et al., who suggested that the differences in response between these two specific types of MOx sensor provides more information on VOC sources [44].

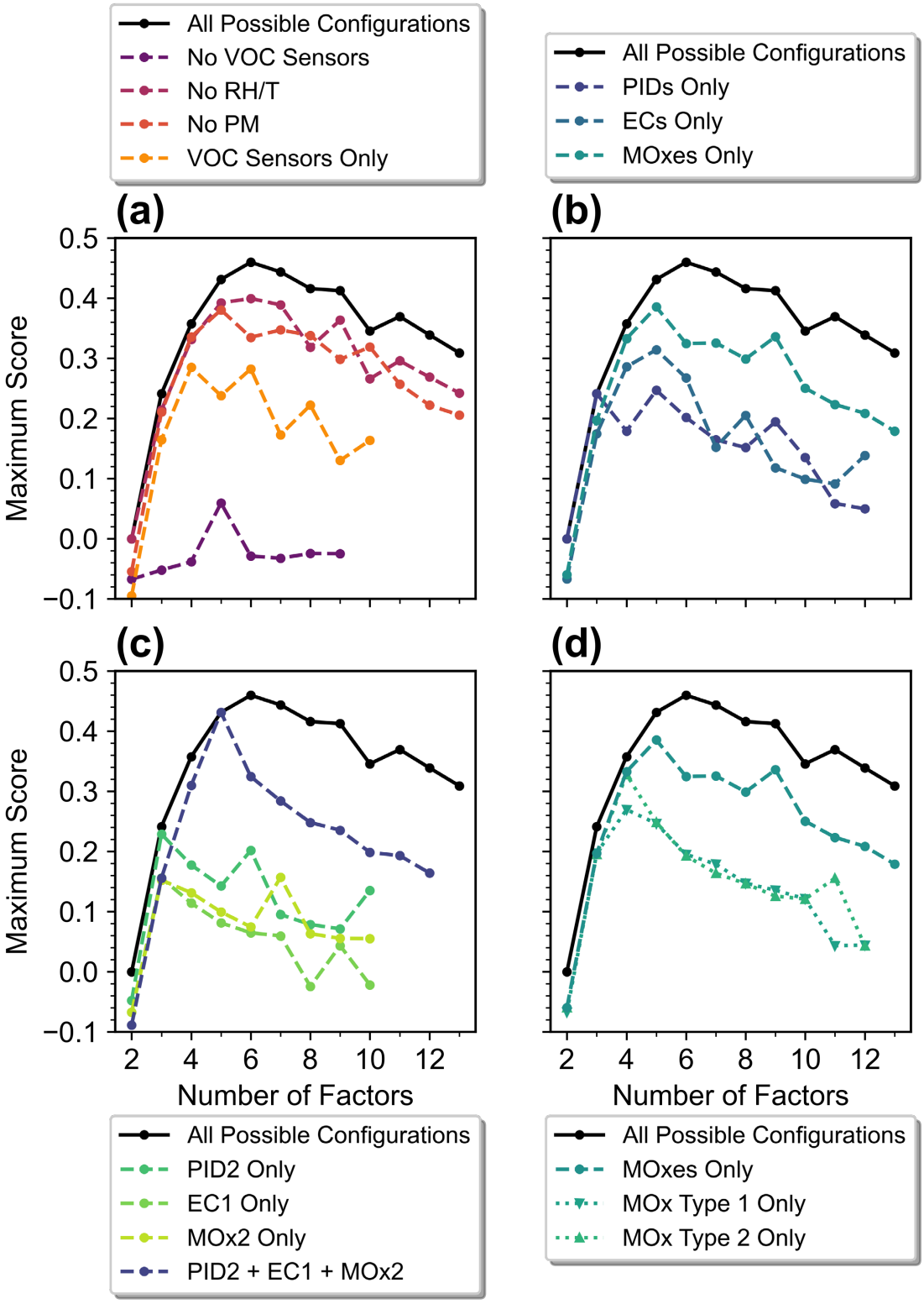

Figure 4.3.4 Comparison of maximum rank-adjusted scores for sensor sub-arrays with various imposed requirements.

4.3.2 Performance Curves for Other Metrics

We explored a few other metrics for scoring the correlations between LCS-derived factors and reference source profiles. Here, we show the rank-adjusted results (e.g. Figure 4.3.4) from three alternative scoring schemes. In the first, we simply adjust the condition in the main text to require correlations of above 0.7. Figure 4.3.5 shows that for this condition, sub-arrays optimized for ranks of k = 4, 6, and 8 have the best performance relative to the number of factors. In the second, we attempt to maximize the sum of LCS correlations with "chemical cocktail," pesticide, and fresh burning factors. Because this condition maximizes information about a small subset of VOC emissions, Figure 4.3.6 shows that the performance curve is maximized at a lower rank of k = 4, and the distribution of curves for different conditions is much narrower. In the third, we maximize the sum of all LCS correlations with "spiky," event-based source profiles (determined by filtering for source profiles with a spectral entropy of $>$ 0.5 [111]), excluding the VOC emission profiles associated with RH-driven surface emissions. Figure 4.3.7 shows that this score is maximized at k = 4 and 6, but unlike the other conditions, there are no sub-arrays with withheld data that approach the maximum performance curve.

Regardless of any quantitative differences across the results of these different metrics, the earlier qualitative conclusions about the relative importance of sensor measurements still hold. Most importantly, we see that the inclusion of multiple VOC sensing technologies results in better performance for all of these alternative metrics, further strengthening our hypothesis about the importance of a multi-dimensional sensor array.

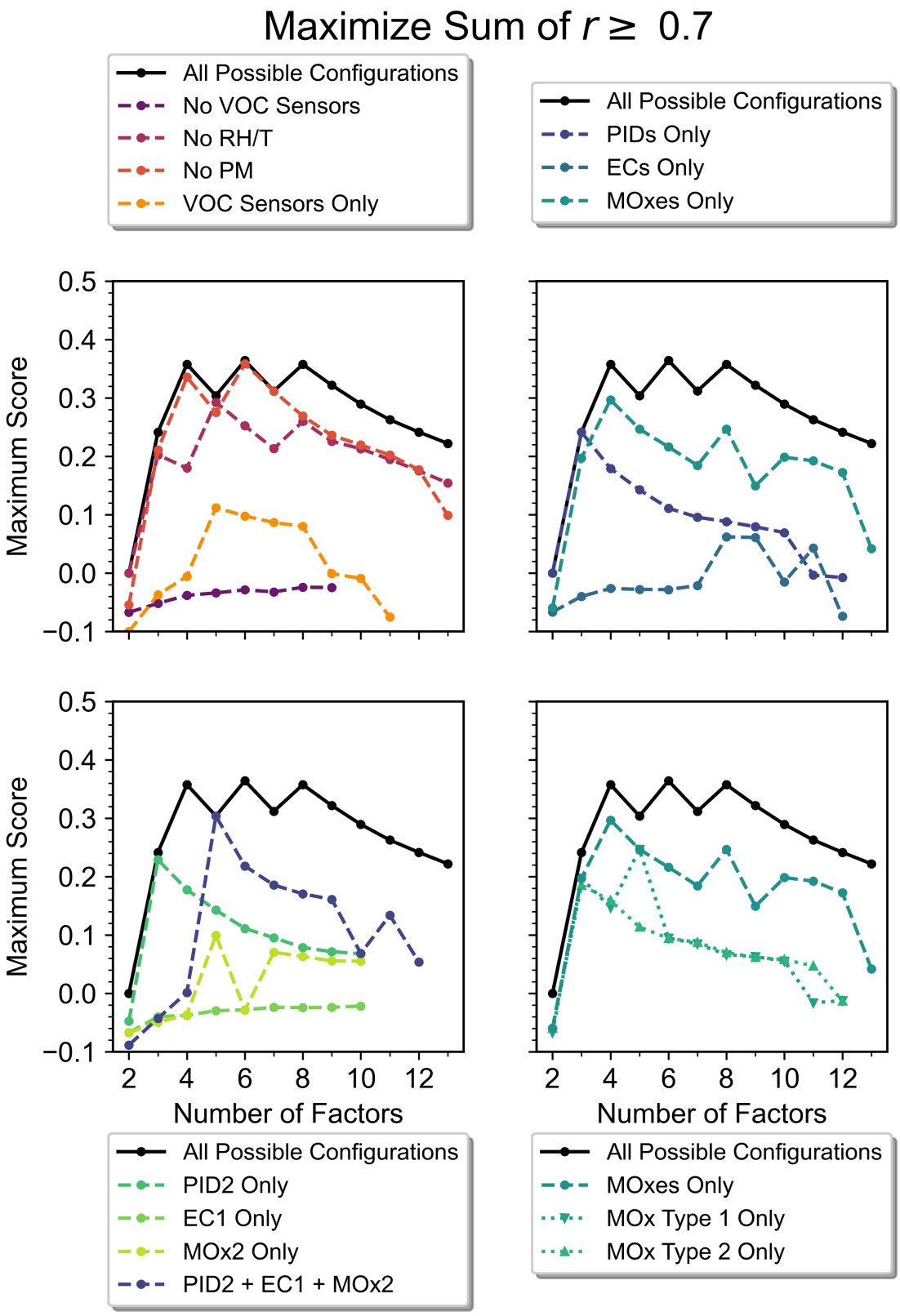

Figure 4.3.5 Comparison of maximum rank-adjusted score (maximize correlations above 0.7) for sensor sub-arrays with various imposed requirements.

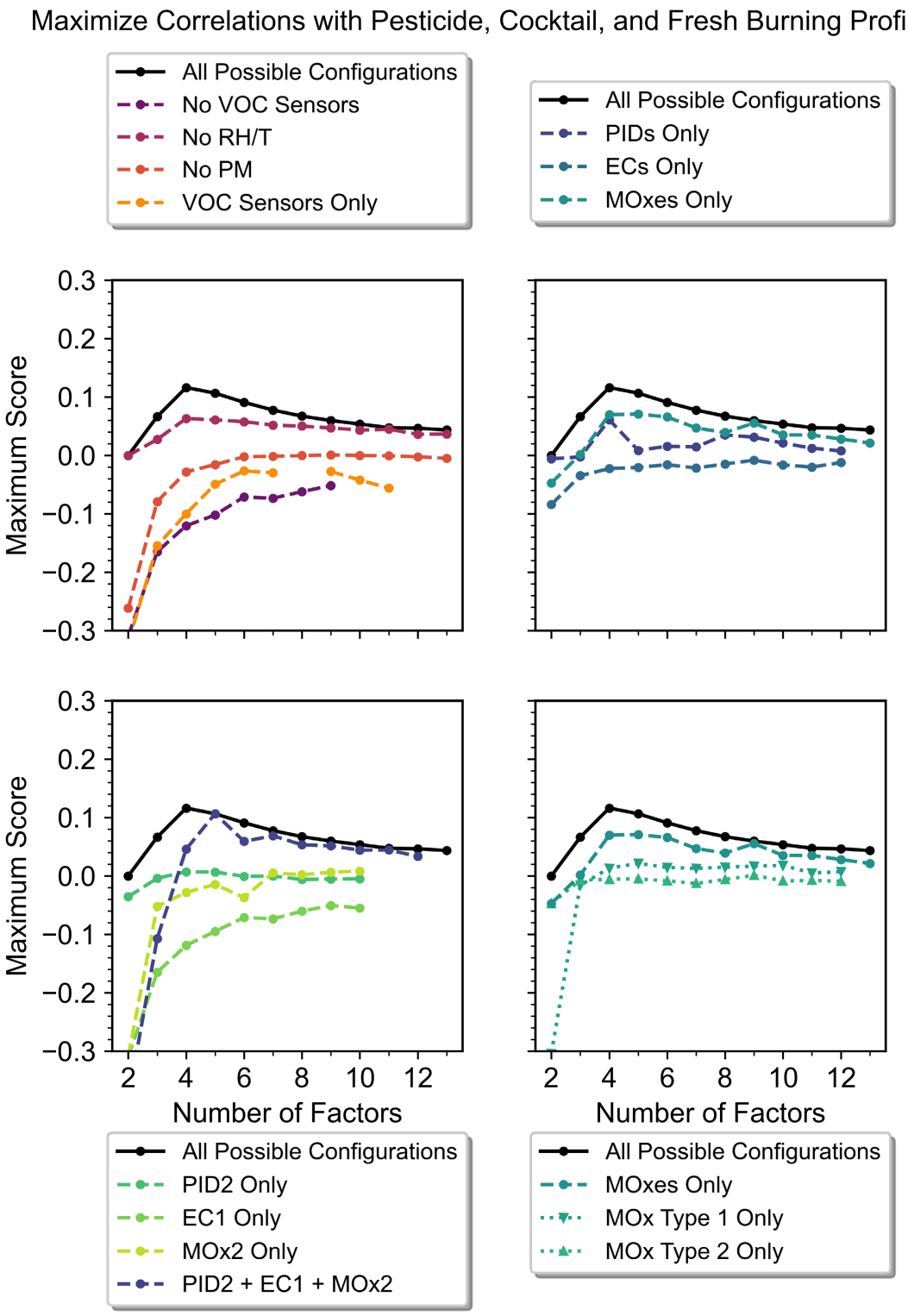

Figure 4.3.6 Comparison of maximum rank-adjusted score (maximize correlations for cocktail, pesticide, and fresh burning profiles) for sensor sub-arrays with various imposed requirements.

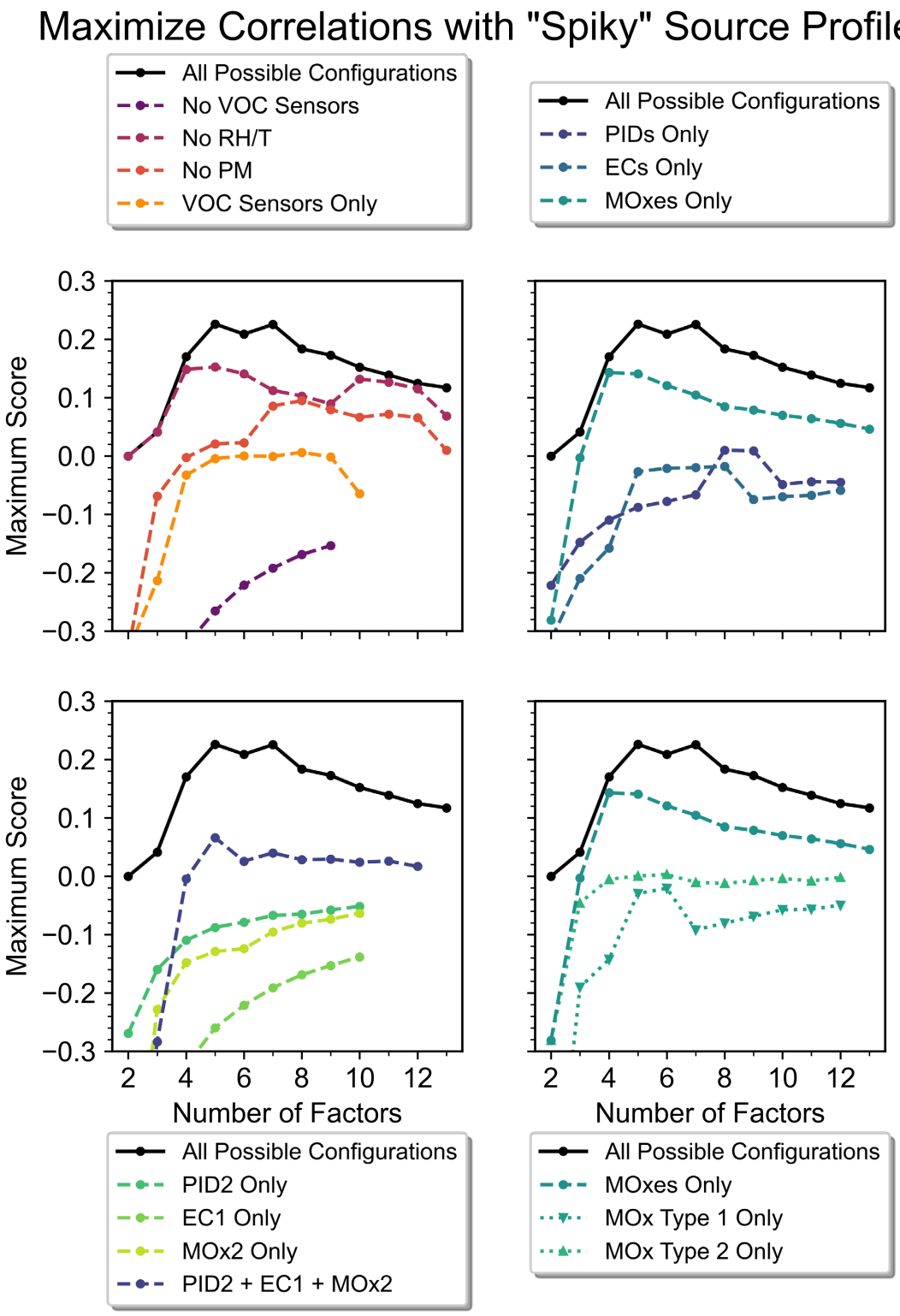

Figure 4.3.7 Comparison of maximum rank-adjusted score (maximize correlations with 'spiky' series) for sensor sub-arrays with various imposed requirements.

4.4 Conclusions and Future Work

In this chapter, we explored the ability of sensor sub-arrays to optimize the NMF analysis used in Chapter 3. We defined a simple scoring function that characterizes the degree to which LCS-derived NMF factors describe known VOC source profiles, with an additional condition imposed to account for the diminishing returns of higher-rank NMF solutions. We conducted an exhaustive, brute-force search of a parameter space that included every possible configuration of low-cost VOC sensors, as well as various combinations of environmental and PM data. The procedure described here could be utilized to assess potential future applications of LCS arrays. With enough *a priori* knowledge about sensor responses and VOC sources, this procedure could even be applied to synthetic sensor data in order to inform the design of future sensor arrays before the process of instrument manufacture even begins.

This analysis of sensor array configurations gives us valuable information about the relative importance of various low-cost measurements in characterizing indoor VOC sources. We have demonstrated that including low-cost PM and environmental measurements alongside VOC sensor measurements helps to characterize multi-pollutant VOC source profiles. In addition, the inclusion of multiple VOC sensing technologies is important for accurately characterizing reference source profiles, even in the most extreme case where only one sensor of each type is included. However, these results also suggest that we should not reduce the sensor array to that extreme: most of the "optimized" sub-arrays found in this analysis include multiple sensors of each measurement type. In section 4.3.2, we show that these observations about optimal sensor sub-array composition are consistent across different choices of scoring function.

The results from this chapter also help to strengthen the conclusions about sensor limitations made in Chapter 3: notably, the results from this exhaustive search of sensor array configurations confirms that LCS-derived NMF factors cannot characterize certain known VOC source profiles, such as cooking and human emissions. However, this sensor array can characterize other VOC source profiles quite well, and the degree to which LCS-derived NMF factors line up with source profiles is partially determined by choice in sensors.

These results allow us to make a few recommendations for future designs of a low-cost measurement that aims to measure indoor VOCs. The first key recommendation is that multiple VOC sensing technologies are included, with an emphasis on the inclusion of PID and MOx sensors. Inclusion of a single EC sensor is helpful but not strictly necessary. The second is that multiple PID sensors and MOx sensors should be included: more specifically, these PID sensors should have different lamp energies or chemical filters and these MOx sensors should represent at least two different types, operated at a variety of heater voltages. Finally, the inclusion of ancillary, non-VOC measurements is important: both PM and environmental measurements help to extract information about indoor VOC sources from the array. We hope that the insights on sensor selection discussed here will help to inform future applications of LCS arrays in characterizing VOC sources.

Appendix

4.A Appendix

4.A.1 Problems with Higher-Rank NMF Solutions

A past application of factor analysis to synthetic environmental data saw factor "splitting" and "mixing" in higher-factor solutions, or cases where a true factor can be mathematically represented by the sum (or another simple matrix tranformation) of multiple factors [101]. The addition of more factors may appear to improve certain performance metrics but "split" or "mixed" factors ultimately fail to explain more of the variability in the dataset. In our analysis, we avoid this artifact of higher factor solutions by prioritizing a rank-scaled metric. Here, we will discuss the optimized sub-array for k = 13 NMF factors, which appears to have high performance in Figure 4.3.1a, but may not actually give much more information about VOC sources than the sub-arrays optimized for lower-dimensional factor analysis. In Figure 4.A.1, we show the correlation matrix between this 13-factor NMF solution and the reference source profiles.

We notice that several of these factors correlate with the same source profiles. To investigate whether these factors are the result of splitting or mixing, we can fit a multiple linear regression (with forced positive coefficients) of LCS factors to the reference source profiles. Regressions for each of the 15 source profiles were cross-validated using the same repeated-holdout procedure described in section 3.A.3, run with 70% of the data used for training and 20% for testing, and in figure 4.A.2 we show the results for MLR models that had acceptable cross-validation statistics (e.g. no over-fitting). The left panel of this figure shows the fitted coefficients of each LCS factor, while the right panel shows the improvement in correlation achieved with the combination of fractional LCS factors compared to the highest correlation achieved by any single factor. We see that there is potentially some factor splitting or mixing at higher rank. For example, the "chemical cocktail" signal appears to be primarily split between LCS factor 10, which is strongly correlated to this source profile, and factor 5, which does not have any strong correlations with any source profiles. Similarly, the pesticide signal is split between factors 12 and 5. Thus, it stands to reason that factor 5 is likely a mathematical artifact of having too many factors, as it's clearly a mix of the true pesticide and cocktail profiles that doesn't give us any new information on indoor VOC sources.

4.A.2 Performance Curves Without Rank Adjustment

In the main text, we plotted the performance curves of various constrained sub-arrays as a function of the rank-scaled metric. In Figure 4.A.3, we show the results of Figure 4.3.4 that are not adjusted with the condition shown in Equation 4.1.

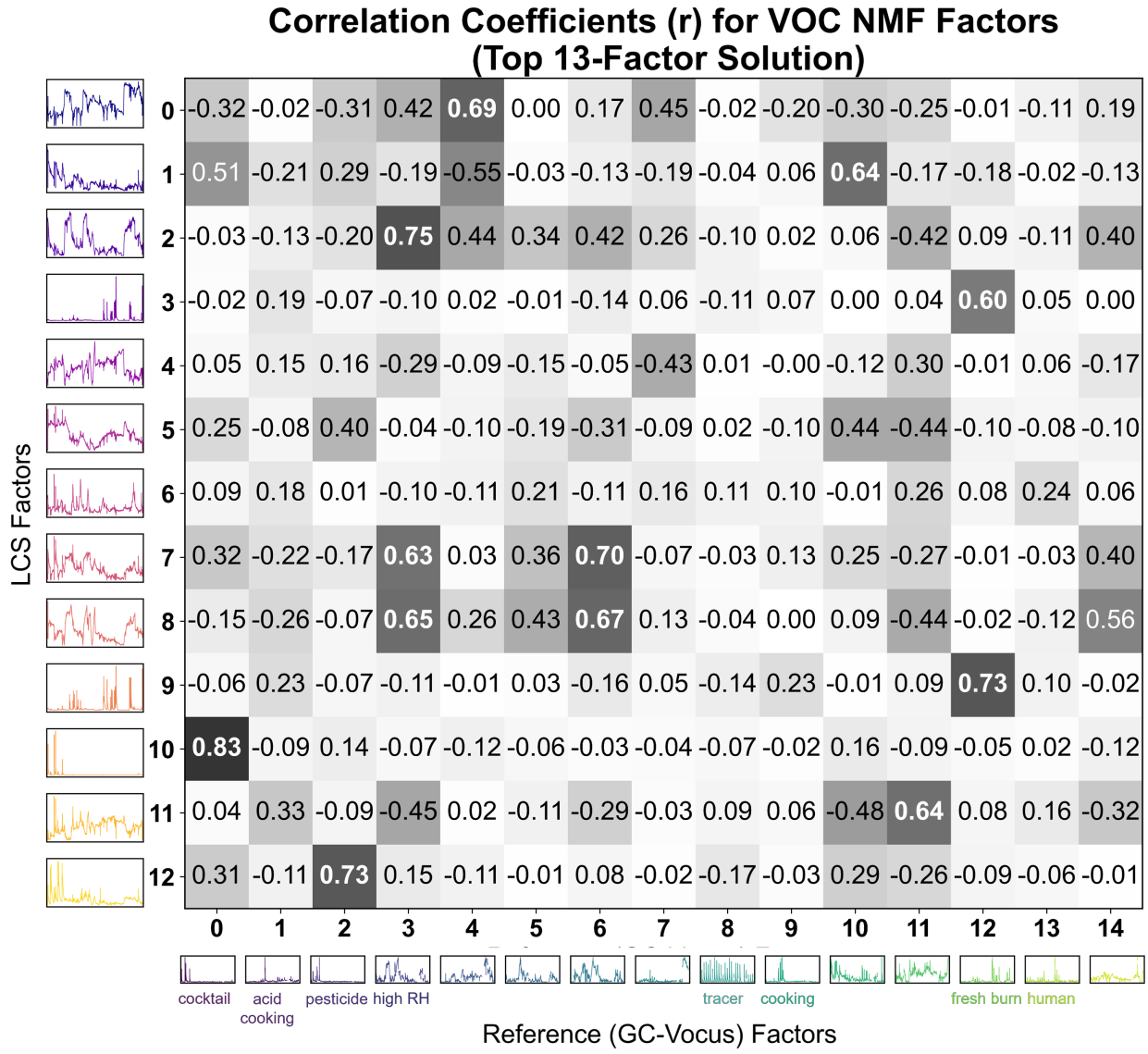

Figure 4.A.1 Correlations between the LCS-derived NMF factors and reference source profiles, for the sensor configuration optimized for a 13-factor solution.

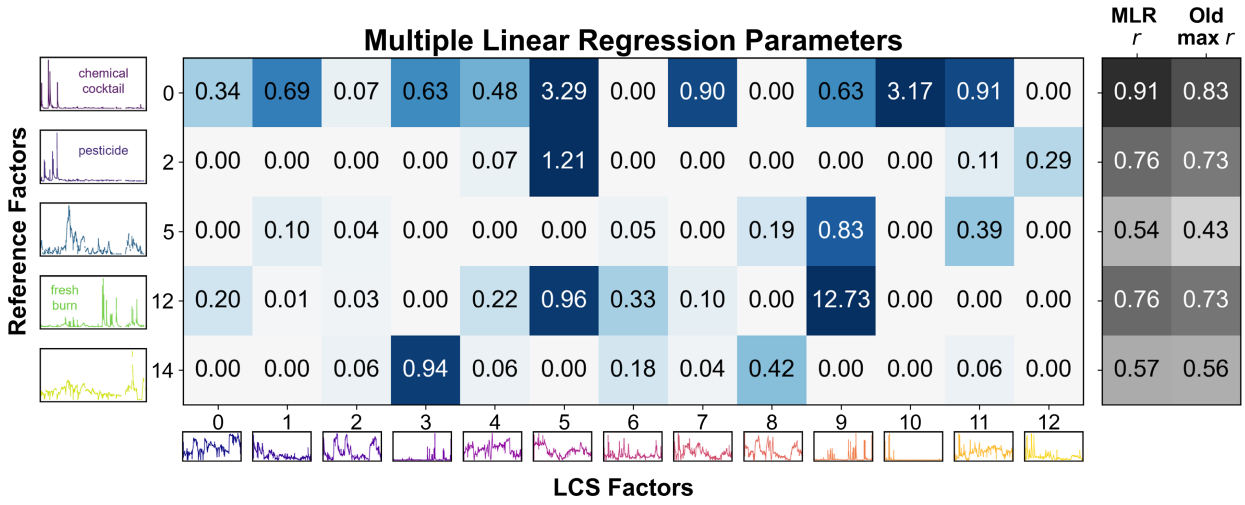

Figure 4.A.2 Multiple linear regression coefficients for LCS factors fit to the corresponding source profile (left); Improvement of the correlation between the MLR model and the corresponding source profile (right).

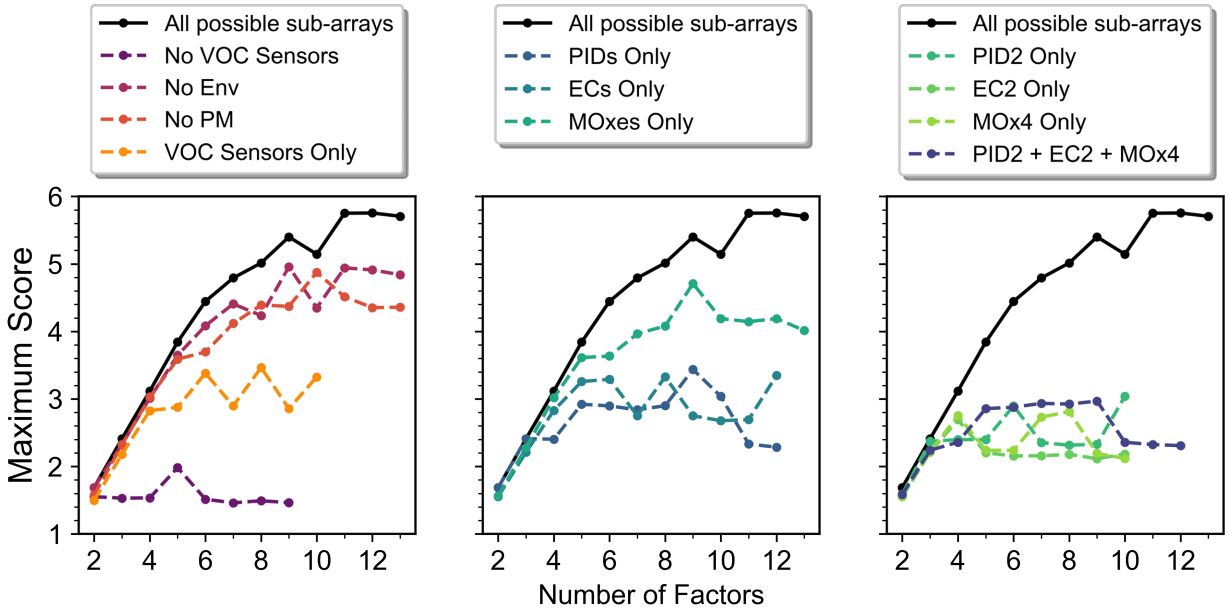

Figure 4.A.3 Comparison of maximum scores for sensor sub-arrays with various imposed requirements, with no NMF rank adjustment.

Chapter 5

Conclusion and Future Directions

5.1 Main Findings

Real-time measurements of atmospheric VOCs improve our understanding of their chemistry and dynamics and of human exposures to these harmful compounds. A feasible low-cost alternative to the costly research-grade instruments that are typically used to make these measurements would open the possibility of widespread, spatially distributed measurements of VOCs in air quality and chemistry contexts. This thesis carries out an in-depth investigation of the suitability of making measurements of environmental VOCs with low-cost sensors, via the development, characterization, optimization, and use of a novel low-cost instrument for measuring environmental VOCs.

In Chapter 2, we described the development and design of the low-cost VOC instrument, which contains an array of low-cost VOC sensors representing three fundamentally different sensor types and takes advantage of user-controlled parameters that achieve greater degrees of differentiation between responses of sensors with the same measurement type. We used this novel instrument to obtain calibration curves for ten typical atmospheric VOCs between 5-100 ppb in a controlled laboratory environment and explored sensor responses to varying RH and binary mixtures, finding that all sensor responses were consistent with results of prior studies and expectations based on physical sensing principles. Our laboratory results demonstrate that this instrument can give quantitative, chemically specific information about VOCs. However, these results also suggest that a truly complete laboratory characterization is not feasible for environmental applications due to the complicated effects on sensor response caused by interaction effects in gas mixtures and variable RH.

In Chapter 3, we described measurements of indoor VOC sources made using our low-cost VOC instrument. Our sensor array and co-located reference monitors made measurements of various chemical perturbations representing realistic sources of VOCs, such as cooking, pesticide application, and wildfire smoke intrusion. Multi-pollutant low-cost measurements, including low-cost VOC measurements, were used to perform a factor analysis that identified periods of time when the air was influenced by different sources of indoor pollution. Results from the LCS-derived factorization were then compared to an independent source apportionment, performed using research-grade reference VOC measurements, that we demonstrated to be highly representative of known VOC sources. This comparison analysis showed that LCS measurements provide meaningful information about several VOC sources caused by activities conducted during the CASA field exper-

iment, and LCS measurements also partially captured the behavior of RH- and temperature-driven changes in VOC emissions from indoor surfaces. Despite limitations related to sensor sensitivities to certain compounds and low emission rates of certain sources, our results show that uncalibrated LCS data provides interpretable insights into the sources and composition of indoor air pollution.

In Chapter 4, we described a procedure for sensor selection to identify any sub-arrays that can provide the same amount of VOC information as the full array. To do this, we evaluated many possible sensor sub-arrays using a "brute force" search process and a simple scoring scheme based on the matrix factorization analysis developed in Chapter 3. The quantitative results from this exhaustive analysis of sensor array configurations confirm a key hypothesis underlying this work: that the inclusion of multiple sensor technologies and operational parameters improves the ability of a sensor array to give useful information about VOC sources and composition. In addition, this sensor selection procedure yielded valuable information about the relative importance of various low-cost measurements in characterizing indoor VOC sources, and our results suggested that the inclusion of ancillary low-cost measurements of environmental dewpoint and PM helps to extract information about indoor VOC sources from the sensor array. Finally, we show that application of this information significantly reduces the size of the LCS array, helping future similar applications to avoid measurement redundancies and minimize material cost.

Overall, the results from this thesis show that while this LCS instrument cannot match the sensitivity or chemical detail expected from research-grade instruments measuring environmental VOCs, it can provide useful, quantitative information about VOC sources and composition at a fraction of the size and cost–opening the possibility of widespread and spatially distributed measurements of VOCs in air quality and chemistry contexts, especially for indoor air.

5.2 Future Work

The instrument we described in this thesis can serve as a prototype for the future development of similar sensor arrays to measure atmospheric VOCs. Our results suggest that there is room for improvement in the methods used for developing and characterizing these future instruments. These suggested improvements would enable the next generation of low-cost VOC instruments to be used in exciting and novel applications. This section describes a few possibilities for future work and applications that would help to improve our understanding of both LCS abilities and VOC composition and sources.

Improving the Transferability of Laboratory Results to the Field

We showed that a complete characterization that builds on the results of Chapter 2 is not feasible: the real atmosphere contains far too many VOCs, with too many RH conditions and possible VOC mixture compositions, for such prescriptive lab characterization to be possible. Instead, we suggest that future work could develop alternative laboratory procedures to characterize sensor responses that are more directly transferable to real-world VOC sources and composition. This could involve a similar approach to the one developed by Thorson et al. [112], which involves calibrating the sensor array with complex mixtures of VOCs representing realistic sources of VOCs, such as emissions of wildfire smoke or cooking. Sensor baseline responses to mixtures representing realistic background air concentrations for various environments (e.g. urban, pristine outdoor, typical indoor)

would also be useful to characterize.

Development of a Physics-Based Model for Low-Cost Sensor Responses to VOCs

A physics-based model that can predict sensor responses as a function of VOC inputs would be extremely useful for optimizing the design and development of a low-cost VOC instrument. There are several detailed studies that elucidate the physical sensing mechanisms of these low-cost sensing technologies [24], [50], [54] and in the last few years, physics-based models have been developed for electrochemical [113] and metal oxide [114] sensors measuring NOx. However, development of physics-based models for low-cost VOC sensors would be challenging, as most manufacturers of commercially available sensors do not disclose important sensor parameters that would greatly affect modeled sensor outputs. Nonetheless, these parameters could potentially be inferred via iterative comparison of model results to laboratory results. A working model of VOC sensor responses that is validated by laboratory results would be invaluable for future LCS applications, as it would allow us gain a very thorough picture of sensor array abilities and limitations.

Low-Cost Measurements of Realistic Outdoor VOCs

Our results from Chapter 3 suggest that LCS are well-suited for measuring sources of VOCs with relatively high concentrations and high variability. There are several outdoor locations that also fit this description, such as areas experiencing wildfire smoke episodes or very polluted urban centers, and measurements of these outdoor VOC sources give valuable insight into the processes driving the secondary formation of ground-level ozone and PM [115]. Future work could attempt to characterize the VOC sources seen at these outdoor sites with a low-cost VOC sensor array and appropriate low-cost ancillary measurements, using a similar approach to the one described in this thesis. However, such outdoor applications will be complicated by relative humidity extremes, which will significantly affect sensor responses and, in cases of prolonged exposure, will even degrade and destroy sensors. Extreme fluctuations in outdoor temperature will not have as great of an effect on sensor responses, but may also pose problems to instrument operation. In order to make low-cost measurements of outdoor VOCs, we would first need to ensure that that a sensor array can operate and provide interpretable data in a wide range of environmental conditions.

As we discussed in Section 2.3.2, one solution to this problem is to pre-treat samples of environmental VOCs by selectively removing water vapor to keep sampled air at a constant and moderate RH level. Such a dryer system must be carefully selected to avoid unintentional removal or introduction of VOCs to the sample stream–for example, Nafion dryers, commonly used to pre-treat measurements of inorganic atmospheric pollutants, have been observed to significantly deplete polar VOCs [74] and certain hydrocarbons [75] in the sample stream. There are a few alternative dryer systems that have been shown to have better VOC recovery rates: Beghi et al. found that diffusion through poly(vinyl fluoride) Tedlar film removes water effectively without causing significant VOC loss [76], and Lee et al. found that drying systems based on Peltier cooling generally preserve polar VOC concentrations [77]. Further work is needed to assess the suitability of these methods for an outdoor, low-cost VOC instrument deployment.

Spatially Distributed Measurements of VOCs

Perhaps the most important feature of low-cost VOC measurements is their potential for spatial distribution, which gives a clear advantage over the more accurate reference-grade monitors. The work described here indicates that LCS perform well in characterizing indoor VOC sources. A natural next step would be to make spatially distributed indoor measurements with multiple VOC sensor nodes at different locations within the building, with at least two of these sensors co-located with reference instruments. These co-located reference measurements could be used to assess the suitability and transferability of the LCS data analysis approach across nodes. Distributed, low-cost VOC measurements could provide valuable and novel information about indoor VOC heterogeneity and transport throughout an indoor space.

Ultimately, we hope that the technologies and approaches described in this work will help the future development of sensor array applications that provide spatially distributed, real-time measurements of VOCs and contribute to our fundamental understanding of chemical composition and human exposure across scales.

References

[1] C. J. L. Murray, A. Y. Aravkin, P. Zheng, *et al.*, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," English, *The Lancet*, vol. 396, no. 10258, pp. 1223–1249, Oct. 2020, Publisher: Elsevier, ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30752-2. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30752-2/fulltext#sec1 (visited on 03/09/2024).

[2] A. Pozzer, S. C. Anenberg, S. Dey, A. Haines, J. Lelieveld, and S. Chowdhury, "Mortality Attributable to Ambient Air Pollution: A Review of Global Estimates," *GeoHealth*, vol. 7, no. 1, e2022GH000711, Jan. 2023, ISSN: 2471-1403. DOI: 10.1029/2022GH000711. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9828848/ (visited on 03/09/2024).

[3] J. T. Pryor, L. O. Cowley, and S. E. Simonds, "The Physiological Effects of Air Pollution: Particulate Matter, Physiology and Disease," *Frontiers in Public Health*, vol. 10, p. 882 569, Jul. 2022, ISSN: 2296-2565. DOI: 10.3389/fpubh.2022.882569. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9329703/ (visited on 03/09/2024).

[4] T. Münzel, T. Gori, S. Al-Kindi, J. Deanfield, J. Lelieveld, A. Daiber, and S. Rajagopalan, "Effects of gaseous and solid constituents of air pollution on endothelial function," en, *European Heart Journal*, vol. 39, no. 38, p. 3543, Oct. 2018, Publisher: Oxford University Press. DOI: 10.1093/eurheartj/ehy481. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6174028/ (visited on 03/09/2024).

[5] A. A. Mitku, T. Zewotir, D. North, P. Jeena, K. Asharam, S. Muttoo, H. Tularam, and R. N. Naidoo, "Impact of ambient air pollution exposure during pregnancy on adverse birth outcomes: Generalized structural equation modeling approach," *BMC Public Health*, vol. 23, no. 1, p. 45, Jan. 2023, ISSN: 1471-2458. DOI: 10.1186/s12889-022-14971-3. [Online]. Available: https://doi.org/10.1186/s12889-022-14971-3 (visited on 03/09/2024).

[6] M. J. Nieuwenhuijsen, P. Dadvand, J. Grellier, D. Martinez, and M. Vrijheid, "Environmental risk factors of pregnancy outcomes: A summary of recent meta-analyses of epidemiological studies," en, *Environmental Health*, vol. 12, no. 1, p. 6, Jan. 2013, ISSN: 1476-069X. DOI: 10.1186/1476-069X-12-6. [Online]. Available: https://doi.org/10.1186/1476-069X-12-6 (visited on 03/09/2024).

[7] M. Camredon, B. Aumont, J. Lee-Taylor, and S. Madronich, "The SOA/VOC/NOx system: An explicit model of secondary organic aerosol formation," en, *Atmos. Chem. Phys.*, p. 12, 2007.

[8] O. US EPA, *Indoor Air Quality*, en, Reports and Assessments, Nov. 2017. [Online]. Available: https://www.epa.gov/report-environment/indoor-air-quality (visited on 03/24/2021).

[9] U. B. Nurmatov, N. Tagiyeva, S. Semple, G. Devereux, and A. Sheikh, "Volatile organic compounds and risk of asthma and allergy: A systematic review," en, *European Respiratory Review*, vol. 24, no. 135, pp. 92–101, Mar. 2015, Publisher: European Respiratory Society Section: Reviews, ISSN: 0905-9180, 1600-0617. DOI: 10.1183/09059180.00000714. [Online]. Available: https://err.ersjournals.com/content/24/135/92 (visited on 02/20/2024).

[10] N. Liu, Z. Bu, W. Liu, *et al.*, "Health effects of exposure to indoor volatile organic compounds from 1980 to 2017: A systematic review and meta-analysis," en, *Indoor Air*, vol. 32, no. 5, e13038, 2022, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ina.13038, ISSN: 1600-0668. DOI: 10.1111/ina.13038. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ina.13038 (visited on 03/11/2024).

[11] GBD 2016 Risk Factors Collaborators, "Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016," eng, *Lancet (London, England)*, vol. 390, no. 10100, pp. 1345–1422, Sep. 2017, ISSN: 1474-547X. DOI: 10.1016/S0140-6736(17)32366-8.

[12] D. M. Lunderberg, P. K. Misztal, Y. Liu, C. Arata, Y. Tian, K. Kristensen, R. J. Weber, W. W. Nazaroff, and A. H. Goldstein, "High-Resolution Exposure Assessment for Volatile Organic Compounds in Two California Residences," *Environmental Science & Technology*, vol. 55, no. 10, pp. 6740–6751, May 2021, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.0c08304. [Online]. Available: https://doi.org/10.1021/acs.est.0c08304 (visited on 09/12/2021).

[13] C. Arata, P. K. Misztal, Y. Tian, D. M. Lunderberg, K. Kristensen, A. Novoselac, M. E. Vance, D. K. Farmer, W. W. Nazaroff, and A. H. Goldstein, "Volatile organic compound emissions during HOMEChem," en, *Indoor Air*, vol. 00, pp. 1–19, Jun. 2021, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ina.12906, ISSN: 1600-0668. DOI: 10.1111/ina.12906. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ina.12906 (visited on 09/12/2021).

[14] D. Y. C. Leung, "Outdoor-indoor air pollution in urban environment: Challenges and opportunity," English, *Frontiers in Environmental Science*, vol. 2, Jan. 2015, Publisher: Frontiers, ISSN: 2296-665X. DOI: 10.3389/fenvs.2014.00069. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2014.00069 (visited on 03/15/2024).

[15] N. J. Nassikas, M. C. McCormack, G. Ewart, *et al.*, "Indoor Air Sources of Outdoor Air Pollution: Health Consequences, Policy, and Recommendations: An Official American Thoracic Society Workshop Report," *Annals of the American Thoracic Society*, vol. 21, no. 3, pp. 365–376, Mar. 2024, Publisher: American Thoracic Society - AJRCCM, ISSN: 2329-6933. DOI: 10.1513/AnnalsATS.202312-1067ST. [Online]. Available: https://www.atsjournals.org/doi/full/10.1513/AnnalsATS.202312-1067ST (visited on 03/15/2024).

[16] R. W. Pinder, J. M. Klopp, G. Kleiman, G. S. W. Hagler, Y. Awe, and S. Terry, "Opportunities and Challenges for Filling the Air Quality Data Gap in Low- and Middle-Income Countries," en, *Atmospheric environment (Oxford, England : 1994)*, vol. 215, Oct. 2019, Publisher: NIH Public Access. DOI: 10.1016/j.atmosenv.2019.06.032. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7887702/ (visited on 03/09/2024).

[17] S. Hossain, W. Che, and A. K.-H. Lau, "Inter- and Intra-Individual Variability of Personal Health Risk of Combined Particle and Gaseous Pollutants across Selected Urban Microenvironments," en, *International Journal of Environmental Research and Public Health*, vol. 19, no. 1, Jan. 2022, Publisher: Multidisciplinary Digital Publishing Institute (MDPI). DOI: 10.3390/ijerph19010565. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8744794/ (visited on 03/09/2024).

[18] A. C. Lewis, J. D. Lee, P. M. Edwards, *et al.*, "Evaluating the performance of low cost chemical sensors for air pollution research," en, *Faraday Discussions*, vol. 189, no. 0, pp. 85–103, Jul. 2016, Publisher: The Royal Society of Chemistry, ISSN: 1364-5498. DOI: 10.1039/C5FD00201J. [Online]. Available: https://pubs.rsc.org/en/content/articlelanding/2016/fd/c5fd00201j (visited on 03/22/2021).

[19] E. G. Snyder, T. H. Watkins, P. A. Solomon, E. D. Thoma, R. W. Williams, G. S. W. Hagler, D. Shelow, D. A. Hindin, V. J. Kilaru, and P. W. Preuss, "The Changing Paradigm of Air Pollution Monitoring," *Environmental Science & Technology*, vol. 47, no. 20, pp. 11 369–11 377, Oct. 2013, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/es4022602. [Online]. Available: https://doi.org/10.1021/es4022602 (visited on 10/30/2020).

[20] F. Mao, K. Khamis, S. Krause, J. Clark, and D. M. Hannah, "Low-Cost Environmental Sensor Networks: Recent Advances and Future Directions," English, *Frontiers in Earth Science*, vol. 7, 2019, Publisher: Frontiers, ISSN: 2296-6463. DOI: 10.3389/feart.2019.00221. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feart.2019.00221/full (visited on 02/04/2021).

[21] S. Xie, J. R. Meeker, L. Perez, *et al.*, "Feasibility and acceptability of monitoring personal air pollution exposure with sensors for asthma self-management," *Asthma Research and Practice*, vol. 7, no. 1, p. 13, Sep. 2021, ISSN: 2054-7064. DOI: 10.1186/s40733-021-00079-9. [Online]. Available: https://doi.org/10.1186/s40733-021-00079-9 (visited on 03/09/2024).

[22] D. H. Hagan, S. Gani, S. Bhandari, K. Patel, G. Habib, J. S. Apte, L. Hildebrandt Ruiz, and J. H. Kroll, "Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India," *Environmental Science & Technology Letters*, vol. 6, no. 8, pp. 467–472, Aug. 2019, Publisher: American Chemical Society. DOI: 10.1021/acs.estlett.9b00393. [Online]. Available: https://doi.org/10.1021/acs.estlett.9b00393 (visited on 10/31/2020).

[23] M. Badura, P. Batog, A. Drzeniecka-Osiadacz, and P. Modzel, *Evaluation of Low-Cost Sensors for Ambient PM2.5 Monitoring*, en, Research Article, ISSN: 1687-725X Pages: e5096540 Publisher: Hindawi Volume: 2018, Oct. 2018. DOI:

https://doi.org/10.1155/2018/5096540. [Online]. Available: https://www.hindawi.com/journals/js/2018/5096540/ (visited on 11/01/2020).

[24] R. Baron and J. Saffell, "Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review," *ACS Sensors*, vol. 2, no. 11, pp. 1553–1566, Nov. 2017, Publisher: American Chemical Society. DOI: 10.1021/acssensors.7b00620. [Online]. Available: https://doi.org/10.1021/acssensors.7b00620 (visited on 03/22/2021).

[25] D. H. Hagan, G. Isaacman-VanWertz, J. P. Franklin, L. M. M. Wallace, B. D. Kocar, C. L. Heald, and J. H. Kroll, "Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments," en, *Copernicus Publications*, Jan. 2018, Accepted: 2018-04-26T19:27:08Z Publisher: Copernicus GmbH, ISSN: 1867-8548. [Online]. Available: https://dspace.mit.edu/handle/1721.1/114971 (visited on 10/31/2020).

[26] L. Spinelle, M. Gerboles, G. Kok, S. Persijn, and T. Sauerwald, "Review of Portable and Low-Cost Sensors for the Ambient Air Monitoring of Benzene and Other Volatile Organic Compounds," *Sensors (Basel, Switzerland)*, vol. 17, no. 7, Jun. 2017, ISSN: 1424-8220. DOI: 10.3390/s17071520. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5539520/ (visited on 03/21/2021).

[27] A. Guenther, C. N. Hewitt, D. Erickson, *et al.*, "A global model of natural volatile organic compound emissions," en, *Journal of Geophysical Research: Atmospheres*, vol. 100, no. D5, pp. 8873–8892, 1995, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/94JD02950, ISSN: 2156-2202. DOI: 10.1029/94JD02950. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1029/94JD02950 (visited on 03/11/2024).

[28] V. Soni, P. Singh, V. Shree, and V. Goel, "Effects of VOCs on Human Health," en, in *Air Pollution and Control*, ser. Energy, Environment, and Sustainability, N. Sharma, A. K. Agarwal, P. Eastwood, T. Gupta, and A. P. Singh, Eds., Singapore: Springer, 2018, pp. 119–142, ISBN: 978-981-10-7185-0. DOI: 10.1007/978-981-10-7185-0_8. [Online]. Available: https://doi.org/10.1007/978-981-10-7185-0_8 (visited on 03/22/2021).

[29] R. Montero-Montoya, R. López-Vargas, and O. Arellano-Aguilar, "Volatile Organic Compounds in Air: Sources, Distribution, Exposure and Associated Illnesses in Children," en, *Annals of Global Health*, vol. 84, no. 2, p. 225, 2018, Publisher: Ubiquity Press. DOI: 10.29024/aogh.910. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6748254/ (visited on 03/11/2024).

[30] K. S. Docherty, W. Wu, Y. B. Lim, and P. J. Ziemann, "Contributions of Organic Peroxides to Secondary Aerosol Formed from Reactions of Monoterpenes with O3," *Environmental Science & Technology*, vol. 39, no. 11, pp. 4049–4059, Jun. 2005, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/es050228s. [Online]. Available: https://doi.org/10.1021/es050228s (visited on 03/20/2021).

[31] G. Churkina, F. Kuik, B. Bonn, A. Lauer, R. Grote, K. Tomiak, and T. M. Butler, “Effect of VOC Emissions from Vegetation on Air Quality in Berlin during a Heatwave,” *Environmental Science & Technology*, vol. 51, no. 11, pp. 6120–6130, Jun. 2017, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.6b06514. [Online]. Available: https://doi.org/10.1021/acs.est.6b06514 (visited on 03/20/2021).

[32] M. Steinbacher, J. Dommen, C. Ammann, C. Spirig, A. Neftel, and A. S. H. Prevot, “Performance characteristics of a proton-transfer-reaction mass spectrometer (PTR-MS) derived from laboratory and field measurements,” en, *International Journal of Mass Spectrometry*, Proton Transfer Reaction Mass Spectrometry, vol. 239, no. 2, pp. 117–128, Dec. 2004, ISSN: 1387-3806. DOI: 10.1016/j.ijms.2004.07.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1387380604003434 (visited on 03/20/2021).

[33] B. Munson, “Chemical Ionization Mass Spectrometry: Theory and Applications,” en, in *Encyclopedia of Analytical Chemistry*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470027318.a6004, American Cancer Society, 2006, ISBN: 978-0-470-02731-8. DOI: 10.1002/9780470027318.a6004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470027318.a6004 (visited on 03/20/2021).

[34] E. Woolfenden, “Monitoring VOCs in Air Using Sorbent Tubes Followed by Thermal Desorption-Capillary GC Analysis: Summary of Data and Practical Guidelines,” *Journal of the Air & Waste Management Association*, vol. 47, no. 1, pp. 20–36, Jan. 1997, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10473289.1997.10464411, ISSN: 1096-2247. DOI: 10.1080/10473289.1997.10464411. [Online]. Available: https://doi.org/10.1080/10473289.1997.10464411 (visited on 03/13/2024).

[35] J. W. Gardner and P. N. Bartlett, “A brief history of electronic noses,” *Sensors and Actuators B: Chemical*, vol. 18, no. 1, pp. 210–211, Mar. 1994, ISSN: 0925-4005. DOI: 10.1016/0925-4005(94)87085-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0925400594870853 (visited on 02/14/2024).

[36] L. Cheng, Q.-H. Meng, A. J. Lilienthal, and P.-F. Qi, “Development of compact electronic noses: A review,” en, *Measurement Science and Technology*, vol. 32, no. 6, p. 062 002, Apr. 2021, Publisher: IOP Publishing, ISSN: 0957-0233. DOI: 10.1088/1361-6501/abef3b. [Online]. Available: https://dx.doi.org/10.1088/1361-6501/abef3b (visited on 03/04/2024).

[37] Y. Yang, B. Liu, J. Hua, T. Yang, Q. Dai, J. Wu, Y. Feng, and P. K. Hopke, “Global review of source apportionment of volatile organic compounds based on highly time-resolved data from 2015 to 2021,” *Environment International*, vol. 165, p. 107 330, Jul. 2022, ISSN: 0160-4120. DOI: 10.1016/j.envint.2022.107330. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412022002574 (visited on 02/26/2024).

[38] S. Luo, Q. Hao, Z. Xu, G. Zhang, Z. Liang, Y. Gou, X. Wang, F. Chen, Y. He, and C. Jiang, “Composition Characteristics of VOCs in the Atmosphere of the Beibei Urban District of Chongqing: Insights from Long-Term Monitoring,” en, *Atmosphere*, vol. 14, no. 9, p. 1452,

Sep. 2023, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2073-4433. DOI: 10.3390/atmos14091452. [Online]. Available: https://www.mdpi.com/2073-4433/14/9/1452 (visited on 03/11/2024).

[39] B. You, W. Zhou, J. Li, Z. Li, and Y. Sun, "A review of indoor Gaseous organic compounds and human chemical Exposure: Insights from Real-time measurements," *Environment International*, vol. 170, p. 107 611, Dec. 2022, ISSN: 0160-4120. DOI: 10.1016/j.envint.2022.107611. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412022005384 (visited on 03/11/2024).

[40] D. Furuta, T. Sayahi, J. Li, B. Wilson, A. A. Presto, and J. Li, "Characterization of inexpensive metal oxide sensor performance for trace methane detection," English, *Atmospheric Measurement Techniques*, vol. 15, no. 17, pp. 5117–5128, Sep. 2022, Publisher: Copernicus GmbH, ISSN: 1867-1381. DOI: 10.5194/amt-15-5117-2022. [Online]. Available: https://amt.copernicus.org/articles/15/5117/2022/ (visited on 03/11/2024).

[41] G. Domènech-Gil, N. T. Duc, J. J. Wikner, J. Eriksson, S. N. Påledal, D. Puglisi, and D. Bastviken, "Electronic Nose for Improved Environmental Methane Monitoring," *Environmental Science & Technology*, vol. 58, no. 1, pp. 352–361, Jan. 2024, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.3c06945. [Online]. Available: https://doi.org/10.1021/acs.est.3c06945 (visited on 03/11/2024).

[42] E. M. Taguem, L. Mennicken, and A.-C. Romain, "Quantile regression with a metal oxide sensors array for methane prediction over a municipal solid waste treatment plant," *Sensors and Actuators B: Chemical*, vol. 334, p. 129 590, May 2021, ISSN: 0925-4005. DOI: 10.1016/j.snb.2021.129590. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400521001581 (visited on 03/11/2024).

[43] S. De Vito, E. Massera, M. Piga, L. Martinotto, and G. Di Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," en, *Sensors and Actuators B: Chemical*, vol. 129, no. 2, pp. 750–757, Feb. 2008, ISSN: 0925-4005. DOI: 10.1016/j.snb.2007.09.060. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400507007691 (visited on 03/21/2021).

[44] A. M. Collier-Oxandale, J. Thorson, H. Halliday, J. Milford, and M. Hannigan, "Understanding the ability of low-cost MOx sensors to quantify ambient VOCs," English, *Atmospheric Measurement Techniques*, vol. 12, no. 3, pp. 1441–1460, Mar. 2019, Publisher: Copernicus GmbH, ISSN: 1867-1381. DOI: 10.5194/amt-12-1441-2019. [Online]. Available: https://amt.copernicus.org/articles/12/1441/2019/ (visited on 03/21/2021).

[45] S. K. Brown, M. R. Sim, M. J. Abramson, and C. N. Gray, "Concentrations of Volatile Organic Compounds in Indoor Air – A Review," en, *Indoor Air*, vol. 4, no. 2, pp. 123–134, 1994, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0668.1994.t01-2-00007.x, ISSN: 1600-0668. DOI: 10.1111/j.1600-0668.1994.t01-2-00007.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0668.1994.t01-2-00007.x (visited on 09/07/2021).

[46] L. Zhang, F. Tian, H. Nie, L. Dang, G. Li, Q. Ye, and C. Kadri, “Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine,” *Sensors and Actuators B: Chemical*, vol. 174, pp. 114–125, Nov. 2012, ISSN: 0925-4005. DOI: 10.1016/j.snb.2012.07.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400512007010 (visited on 03/13/2024).

[47] E. J. Wolfrum, R. M. Meglen, D. Peterson, and J. Sluiter, “Metal oxide sensor arrays for the detection, differentiation, and quantification of volatile organic compounds at sub-parts-per-million concentration levels,” en, *Sensors and Actuators B: Chemical*, vol. 115, no. 1, pp. 322–329, May 2006, ISSN: 0925-4005. DOI: 10.1016/j.snb.2005.09.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400505008130 (visited on 03/21/2021).

[48] C. Arnold, M. Harms, and J. Goschnick, “Air quality monitoring and fire detection with the Karlsruhe electronic micronose KAMINA,” *IEEE Sensors Journal*, vol. 2, no. 3, pp. 179–188, Jun. 2002, Conference Name: IEEE Sensors Journal, ISSN: 1558-1748. DOI: 10.1109/JSEN.2002.800681. [Online]. Available: https://ieeexplore.ieee.org/document/1021059 (visited on 02/22/2024).

[49] K. Persaud and G. Dodd, “Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose,” en, *Nature*, vol. 299, no. 5881, pp. 352–355, Sep. 1982, Number: 5881 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/299352a0. [Online]. Available: https://www.nature.com/articles/299352a0 (visited on 02/15/2024).

[50] N. Barsan and U. Weimar, “Conduction Model of Metal Oxide Gas Sensors,” en, *Journal of Electroceramics*, vol. 7, no. 3, pp. 143–167, Dec. 2001, ISSN: 1573-8663. DOI: 10.1023/A:1014405811371. [Online]. Available: https://doi.org/10.1023/A:1014405811371 (visited on 02/15/2024).

[51] H. Liu, R. Wu, Q. Guo, Z. Hua, and Y. Wu, “Electronic Nose Based on Temperature Modulation of MOS Sensors for Recognition of Excessive Methanol in Liquors,” *ACS Omega*, vol. 6, no. 45, pp. 30 598–30 606, Nov. 2021, Publisher: American Chemical Society. DOI: 10.1021/acsomega.1c04350. [Online]. Available: https://doi.org/10.1021/acsomega.1c04350 (visited on 03/12/2024).

[52] J. E. Lovelock, “A Photoionization Detector for Gases and Vapours,” en, *Nature*, vol. 188, no. 4748, pp. 401–401, Oct. 1960, Number: 4748 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/188401a0. [Online]. Available: https://www.nature.com/articles/188401a0 (visited on 03/22/2021).

[53] M. Liess and M. Leonhardt, “New operation principle for ultra-stable photo-ionization detectors,” en, *Measurement Science and Technology*, vol. 14, no. 4, pp. 427–432, Apr. 2003, ISSN: 0957-0233. DOI: 10.1088/0957-0233/14/4/304. [Online]. Available: https://iopscience.iop.org/article/10.1088/0957-0233/14/4/304 (visited on 02/23/2023).

[54] A. N. Freedman, “The photoionization detector: Theory, performance and application as a low-level monitor of oil vapour,” *Journal of Chromatography A*, vol. 190, no. 2, pp. 263–273, Apr. 1980, ISSN: 0021-9673. DOI: 10.1016/S0021-9673(00)88229-1. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021967300882291 (visited on 02/19/2024).

[55] C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao, “Metal Oxide Gas Sensors: Sensitivity and Influencing Factors,” *Sensors (Basel, Switzerland)*, vol. 10, no. 3, pp. 2088–2106, Mar. 2010, ISSN: 1424-8220. DOI: 10.3390/s100302088. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3264469/ (visited on 03/22/2021).

[56] P. Wei, Z. Ning, S. Ye, L. Sun, F. Yang, K. C. Wong, D. Westerdahl, and P. K. K. Louie, “Impact Analysis of Temperature and Humidity Conditions on Electrochemical Sensor Response in Ambient Air Quality Monitoring,” en, *Sensors*, vol. 18, no. 2, p. 59, Feb. 2018, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s18020059. [Online]. Available: https://www.mdpi.com/1424-8220/18/2/59 (visited on 03/01/2024).

[57] G. Müller and G. Sberveglieri, “Origin of Baseline Drift in Metal Oxide Gas Sensors: Effects of Bulk Equilibration,” en, *Chemosensors*, vol. 10, no. 5, p. 171, May 2022, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2227-9040. DOI: 10.3390/chemosensors10050171. [Online]. Available: https://www.mdpi.com/2227-9040/10/5/171 (visited on 03/13/2024).

[58] C. Lieber and A. Mahadevan-Jansen, “Automated Method for Subtraction of Fluorescence from Biological Raman Spectra,” *Applied spectroscopy*, vol. 57, pp. 1363–7, Dec. 2003. DOI: 10.1366/000370203322554518.

[59] J. Zhao, H. Lui, D. I. McLean, and H. Zeng, “Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy,” en, *Applied Spectroscopy*, vol. 61, no. 11, pp. 1225–1232, Nov. 2007, Publisher: SAGE Publications Ltd STM, ISSN: 0003-7028. DOI: 10.1366/000370207782597003. [Online]. Available: https://doi.org/10.1366/000370207782597003 (visited on 02/17/2024).

[60] *TGS 2600 Product Information*, Sep. 2013. [Online]. Available: https://www.figarosensor.com/product/docs/TGS2600B00%20%280913%29.pdf (visited on 03/01/2024).

[61] *TGS 2602 Product Information*, Jun. 2015. [Online]. Available: https://www.figarosensor.com/product/docs/TGS2602-B00%20%280615%29.pdf (visited on 03/01/2024).

[62] M. L. Hitchman and J. R. Saffell, “Considerations of Thermodynamics and Kinetics for the Effects of Relative Humidity on the Electrolyte in Electrochemical Toxic Gas Sensors,” *ACS Sensors*, vol. 6, no. 11, pp. 3985–3993, Nov. 2021, Publisher: American Chemical Society. DOI: 10.1021/acssensors.1c01339. [Online]. Available: https://doi.org/10.1021/acssensors.1c01339 (visited on 02/24/2023).

[63] T. V. Adamia, V. L. Budovich, I. A. Nevjagskaya, A. F. Shlyakhov, and S. A. Jakovlev, "Effect of temperature on the sensitivity of the photoionization detector," *Journal of Chromatography A*, vol. 540, pp. 441–448, Jan. 1991, ISSN: 0021-9673. DOI: 10.1016/S0021-9673(01)88836-1. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021967301888361 (visited on 03/13/2024).

[64] A. Scott, *The effects of high humidity on standard PID sensor can result in errors*, en-US, Aug. 2020. [Online]. Available: https://ionscience.com/usa/news/effects-of-high-humidity-on-photoionisation-detectors-pids/ (visited on 02/28/2023).

[65] A. K. Farquhar, G. S. Henshaw, and D. E. Williams, "Understanding and Correcting Unwanted Influences on the Signal from Electrochemical Gas Sensors," *ACS Sensors*, vol. 6, no. 3, pp. 1295–1304, Mar. 2021, Publisher: American Chemical Society. DOI: 10.1021/acssensors.0c02589. [Online]. Available: https://doi.org/10.1021/acssensors.0c02589 (visited on 02/27/2023).

[66] G. F. Fine, L. M. Cavanagh, A. Afonja, and R. Binions, "Metal Oxide Semi-Conductor Gas Sensors in Environmental Monitoring," en, *Sensors*, vol. 10, no. 6, pp. 5469–5502, Jun. 2010, Number: 6 Publisher: Molecular Diversity Preservation International, ISSN: 1424-8220. DOI: 10.3390/s100605469. [Online]. Available: https://www.mdpi.com/1424-8220/10/6/5469 (visited on 02/28/2023).

[67] N. Bârsan and U. Weimar, "Understanding the fundamental principles of metal oxide based gas sensors; the example of CO sensing with SnO2 sensors in the presence of humidity," en, *Journal of Physics: Condensed Matter*, vol. 15, no. 20, R813, May 2003, ISSN: 0953-8984. DOI: 10.1088/0953-8984/15/20/201. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/15/20/201 (visited on 02/28/2023).

[68] G. Korotcenkov, I. Blinov, V. Brinzari, and J. R. Stetter, "Effect of air humidity on gas response of SnO2 thin film ozone sensors," en, *Sensors and Actuators B: Chemical*, vol. 122, no. 2, pp. 519–526, Mar. 2007, ISSN: 0925-4005. DOI: 10.1016/j.snb.2006.06.025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092540050600459X (visited on 02/28/2023).

[69] G. Chabanis, I. P. Parkin, and D. E. Williams, "A simple equivalent circuit model to represent microstructure effects on the response of semiconducting oxide-based gas sensors," *Measurement Science and Technology*, vol. 14, no. 1, pp. 76–86, Jan. 2003, ISSN: 0957-0233. DOI: 10.1088/0957-0233/14/1/312. [Online]. Available: https://iopscience.iop.org/article/10.1088/0957-0233/14/1/312 (visited on 02/19/2024).

[70] G. Chabanis, I. P. Parkin, and D. E. Williams, "A simple equivalent circuit model to represent microstructure effects on the response of semiconducting oxide-based gas sensors," en, *Measurement Science and Technology*, vol. 14, no. 1, pp. 76–86, Jan. 2003, ISSN: 0957-0233. DOI: 10.1088/0957-0233/14/1/312. [Online]. Available: https://iopscience.iop.org/article/10.1088/0957-0233/14/1/312 (visited on 02/19/2024).

[71] J. D. Pleil, K. D. Oliver, and W. A. McClenny, “Enhanced Performance of Nafion Dryers in Removing Water from Air Samples Prior to Gas Chromatographic Analysis,” en, *JAPCA*, vol. 37, no. 3, pp. 244–248, Mar. 1987, ISSN: 0894-0630. DOI: 10.1080/08940630.1987.10466219. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/08940630.1987.10466219 (visited on 04/10/2024).

[72] C. Xiang, W. Wang, X. Zhao, N. Yang, and C. Tu, “Dehumidification performance of the Nafion tube for SO2 at low concentration,” en, *IOP Conference Series: Materials Science and Engineering*, vol. 490, p. 022 036, Apr. 2019, ISSN: 1757-899X. DOI: 10.1088/1757-899X/490/2/022036. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/490/2/022036 (visited on 04/10/2024).

[73] D.-J. Kim, T.-V. Dinh, J.-Y. Lee, D.-J. Son, and J.-C. Kim, “Effect of Nafion Dryer and Cooler on Ambient Air Pollutant (O3, SO2, CO) Measurement,” en, *Asian Journal of Atmospheric Environment*, vol. 14, no. 1, pp. 28–34, Mar. 2020, ISSN: 2287-1160. DOI: 10.5572/ajae.2020.14.1.028. [Online]. Available: https://doi.org/10.5572/ajae.2020.14.1.028 (visited on 04/10/2024).

[74] B. L. Deming, D. Pagonis, X. Liu, D. A. Day, R. Talukdar, J. E. Krechmer, J. A. de Gouw, J. L. Jimenez, and P. J. Ziemann, “Measurements of delays of gas-phase compounds in a wide variety of tubing materials due to gas–wall interactions,” English, *Atmospheric Measurement Techniques*, vol. 12, no. 6, pp. 3453–3461, Jun. 2019, Publisher: Copernicus GmbH, ISSN: 1867-1381. DOI: 10.5194/amt-12-3453-2019. [Online]. Available: https://amt.copernicus.org/articles/12/3453/2019/ (visited on 04/10/2024).

[75] Q. Gong and K. L. Demerjian, “Hydrocarbon Losses on a Regenerated Nation® Dryer,” *Journal of the Air & Waste Management Association*, vol. 45, no. 6, pp. 490–493, Jun. 1995, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10473289.1995.10467379, ISSN: 1096-2247. DOI: 10.1080/10473289.1995.10467379. [Online]. Available: https://doi.org/10.1080/10473289.1995.10467379 (visited on 04/10/2024).

[76] S. Beghi and J.-M. Guillot, “Sample water removal method in volatile organic compound analysis based on diffusion through poly(vinyl fluoride) film,” *Journal of Chromatography A*, vol. 1127, no. 1, pp. 1–5, Sep. 2006, ISSN: 0021-9673. DOI: 10.1016/j.chroma.2006.05.102. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021967306010983 (visited on 04/10/2024).

[77] J.-Y. Lee, T.-V. Dinh, D.-J. Kim, I.-Y. Choi, J.-W. Ahn, S.-Y. Park, Y.-J. Jung, and J.-C. Kim, “Comparison of Water Pretreatment Devices for the Measurement of Polar Odorous Compounds,” en, *Applied Sciences*, vol. 9, no. 19, p. 4045, Jan. 2019, Number: 19 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2076-3417. DOI: 10.3390/app9194045. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/4045 (visited on 04/10/2024).

[78] S.-W. Lee, T.-V. Dinh, S.-Y. Park, I.-Y. Choi, I.-Y. Kim, B.-G. Park, D.-H. Baek, J.-H. Park, Y.-B. Seo, and J.-C. Kim, "Development of a Moisture Pretreatment Device for the Accurate Quantitation of Water-Soluble Volatile Organic Compounds in Air," en, *Chemosensors*, vol. 11, no. 3, p. 188, Mar. 2023, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2227-9040. DOI: 10.3390/chemosensors11030188. [Online]. Available: https://www.mdpi.com/2227-9040/11/3/188 (visited on 04/10/2024).

[79] D. E. Williams and K. F. E. Pratt, "Theory of self-diagnostic sensor array devices using gas-sensitive resistors," en, *Journal of the Chemical Society, Faraday Transactions*, vol. 91, no. 13, pp. 1961–1966, Jan. 1995, Publisher: The Royal Society of Chemistry, ISSN: 1364-5455. DOI: 10.1039/FT9959101961. [Online]. Available: https://pubs.rsc.org/en/content/articlelanding/1995/ft/ft9959101961 (visited on 02/19/2024).

[80] E. Llobet, X. Vilanova, J. Brezmes, J. E. Sueiras, R. Alcubilla, and X. Correig, "Steady-State and Transient Behavior of Thick-Film Tin Oxide Sensors in the Presence of Gas Mixtures," en, *Journal of The Electrochemical Society*, vol. 145, no. 5, p. 1772, May 1998, Publisher: IOP Publishing, ISSN: 1945-7111. DOI: 10.1149/1.1838556. [Online]. Available: https://iopscience.iop.org/article/10.1149/1.1838556/meta (visited on 02/19/2024).

[81] N. Wang, L. Ernle, G. Bekö, P. Wargocki, and J. Williams, "Emission Rates of Volatile Organic Compounds from Humans," *Environmental Science & Technology*, vol. 56, no. 8, pp. 4838–4848, Apr. 2022, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.1c08764. [Online]. Available: https://doi.org/10.1021/acs.est.1c08764 (visited on 03/13/2024).

[82] J. E. Castellini, C. A. Faulkner, W. Zuo, and M. D. Sohn, "Quantifying spatiotemporal variability in occupant exposure to an indoor airborne contaminant with an uncertain source location," en, *Building Simulation*, vol. 16, no. 6, pp. 889–913, Jun. 2023, ISSN: 1996-8744. DOI: 10.1007/s12273-022-0971-3. [Online]. Available: https://doi.org/10.1007/s12273-022-0971-3 (visited on 02/22/2024).

[83] T. Liu, L. Guo, M. Wang, C. Su, D. Wang, H. Dong, J. Chen, and W. Wu, "Review on Algorithm Design in Electronic Noses: Challenges, Status, and Trends," *Intelligent Computing*, vol. 2, p. 0012, Jan. 2023, Publisher: American Association for the Advancement of Science. DOI: 10.34133/icomputing.0012. [Online]. Available: https://spj.science.org/doi/10.34133/icomputing.0012 (visited on 02/22/2024).

[84] J. Krechmer, F. Lopez-Hilfiker, A. Koss, *et al.*, "Evaluation of a New Reagent-Ion Source and Focusing Ion–Molecule Reactor for Use in Proton-Transfer-Reaction Mass Spectrometry," *Analytical Chemistry*, vol. 90, no. 20, pp. 12 011–12 018, Oct. 2018, Publisher: American Chemical Society, ISSN: 0003-2700. DOI: 10.1021/acs.analchem.8b02641. [Online]. Available: https://doi.org/10.1021/acs.analchem.8b02641 (visited on 04/27/2020).

[85] C. Warneke, J. A. de Gouw, W. C. Kuster, P. D. Goldan, and R. Fall, "Validation of Atmospheric VOC Measurements by Proton-Transfer- Reaction Mass Spectrometry Using a Gas-Chromatographic Preseparation Method," *Environmental Science & Technology*,

vol. 37, no. 11, pp. 2494–2501, Jun. 2003, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/es026266i. [Online]. Available: https://doi.org/10.1021/es026266i (visited on 02/22/2024).

[86] *QuantAQ*. [Online]. Available: https://www.quant-aq.com/products/modulair (visited on 02/22/2024).

[87] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," en, *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999, Number: 6755 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/44565. [Online]. Available: https://www.nature.com/articles/44565 (visited on 02/23/2024).

[88] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, ISSN: 1533-7928. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html (visited on 02/28/2024).

[89] B. H. Lee, F. D. Lopez-Hilfiker, E. L. D'Ambro, P. Zhou, M. Boy, T. Petäjä, L. Hao, A. Virtanen, and J. A. Thornton, "Semi-volatile and highly oxygenated gaseous and particulate organic compounds observed above a boreal forest canopy," English, *Atmospheric Chemistry and Physics*, vol. 18, no. 15, pp. 11 547–11 562, Aug. 2018, Publisher: Copernicus GmbH, ISSN: 1680-7316. DOI: 10.5194/acp-18-11547-2018. [Online]. Available: https://acp.copernicus.org/articles/18/11547/2018/ (visited on 02/23/2024).

[90] P. K. Hopke, *A Guide to Positive Matrix Factorization*. [Online]. Available: https://people.clarkson.edu/~phopke/PMF-Guidance.htm (visited on 02/23/2024).

[91] D. H. Hagan and J. H. Kroll, "Assessing the accuracy of low-cost optical particle sensors using a physics-based approach," en, *Atmospheric Measurement Techniques*, vol. 13, no. 11, pp. 6343–6355, Nov. 2020, ISSN: 1867-8548. DOI: 10.5194/amt-13-6343-2020. [Online]. Available: https://amt.copernicus.org/articles/13/6343/2020/ (visited on 05/20/2021).

[92] A. B. Owen and P. O. Perry, "Bi-cross-validation of the SVD and the nonnegative matrix factorization," *The Annals of Applied Statistics*, vol. 3, no. 2, pp. 564–594, Jun. 2009, Publisher: Institute of Mathematical Statistics, ISSN: 1932-6157, 1941-7330. DOI: 10.1214/08-AOAS227. [Online]. Available: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-3/issue-2/Bi-cross-validation-of-the-SVD-and-the-nonnegative-matrix/10.1214/08-AOAS227.full (visited on 02/23/2024).

[93] X. Lin and P. C. Boutros, "Optimization and expansion of non-negative matrix factorization," *BMC Bioinformatics*, vol. 21, no. 1, p. 7, Jan. 2020, ISSN: 1471-2105. DOI: 10.1186/s12859-019-3312-5. [Online]. Available: https://doi.org/10.1186/s12859-019-3312-5 (visited on 02/23/2024).

[94] G. Gilad, I. Sason, and R. Sharan, “An automated approach for determining the number of components in non-negative matrix factorization with application to mutational signature learning,” en, *Machine Learning: Science and Technology*, vol. 2, no. 1, p. 015 013, Mar. 2021, ISSN: 2632-2153. DOI: 10.1088/2632-2153/abc60a. [Online]. Available: https://iopscience.iop.org/article/10.1088/2632-2153/abc60a (visited on 02/23/2024).

[95] H. Li, T. G. Almeida, Y. Luo, *et al.*, “Fragmentation inside proton-transfer-reaction-based mass spectrometers limits the detection of ROOR and ROOH peroxides,” English, *Atmospheric Measurement Techniques*, vol. 15, no. 6, pp. 1811–1827, Mar. 2022, Publisher: Copernicus GmbH, ISSN: 1867-1381. DOI: 10.5194/amt-15-1811-2022. [Online]. Available: https://amt.copernicus.org/articles/15/1811/2022/ (visited on 03/13/2024).

[96] F. Li, D. D. Huang, L. Tian, *et al.*, “Response of protonated, adduct, and fragmented ions in Vocus proton-transfer-reaction time-of-flight mass spectrometer (PTR-ToF-MS),” English, *EGUsphere*, pp. 1–27, Jan. 2024, Publisher: Copernicus GmbH. DOI: 10.5194/egusphere-2024-16. [Online]. Available: https://egusphere.copernicus.org/preprints/2024/egusphere-2024-16/ (visited on 02/28/2024).

[97] C. Zhou, Y. Zhan, S. Chen, M. Xia, C. Ronda, M. Sun, H. Chen, and X. Shen, “Combined effects of temperature and humidity on indoor VOCs pollution: Intercity comparison,” *Building and Environment*, vol. 121, pp. 26–34, Aug. 2017, ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2017.04.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360132317301737 (visited on 02/28/2024).

[98] S. Zhou, H. Liu, Y. Ding, and Y. Wu, “The effects of temperature and humidity on the VOC emission rate from dry building materials,” en, *IOP Conference Series: Materials Science and Engineering*, vol. 609, no. 4, p. 042 001, Sep. 2019, ISSN: 1757-8981, 1757-899X. DOI: 10.1088/1757-899X/609/4/042001. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/609/4/042001 (visited on 02/28/2024).

[99] P. Markowicz and L. Larsson, “Influence of relative humidity on VOC concentrations in indoor air,” en, *Environmental Science and Pollution Research*, vol. 22, no. 8, pp. 5772–5779, Apr. 2015, ISSN: 1614-7499. DOI: 10.1007/s11356-014-3678-x. [Online]. Available: https://doi.org/10.1007/s11356-014-3678-x (visited on 02/28/2024).

[100] S. Hjerrild Smedemark, M. Ryhl-Svendsen, and A. Schieweck, “Quantification of formic acid and acetic acid emissions from heritage collections under indoor room conditions. Part I: Laboratory and field measurements,” *Heritage Science*, vol. 8, no. 58, 2020, ISSN: 2050-7445. DOI: 10.1186/s40494-020-00404-0.

[101] I. M. Ulbrich, M. R. Canagaratna, Q. Zhang, D. R. Worsnop, and J. L. Jimenez, “Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data,” English, *Atmospheric Chemistry and Physics*, vol. 9, no. 9, pp. 2891–2918, May 2009, Publisher: Copernicus GmbH, ISSN: 1680-7316. DOI: 10.5194/acp-9-2891-2009. [Online]. Available: https://acp.copernicus.org/articles/9/2891/2009/acp-9-2891-2009.html (visited on 03/06/2024).

[102] D. M. Lunderberg, K. Kristensen, Y. Tian, *et al.*, "Surface Emissions Modulate Indoor SVOC Concentrations through Volatility-Dependent Partitioning," *Environmental Science & Technology*, vol. 54, no. 11, pp. 6751–6760, Jun. 2020, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.0c00966. [Online]. Available: https://doi.org/10.1021/acs.est.0c00966 (visited on 04/11/2024).

[103] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, Mar. 1947, Publisher: Institute of Mathematical Statistics, ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177730491. [Online]. Available: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-18/issue-1/On-a-Test-of-Whether-one-of-Two-Random-Variables/10.1214/aoms/1177730491.full (visited on 03/14/2024).

[104] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," en, *Machine Learning*, vol. 109, no. 11, pp. 1997–2028, Nov. 2020, ISSN: 1573-0565. DOI: 10.1007/s10994-020-05910-7. [Online]. Available: https://doi.org/10.1007/s10994-020-05910-7 (visited on 02/26/2024).

[105] R. G. Buttery, L. C. Ling, and D. J. Stern, "Studies on Popcorn Aroma and Flavor Volatiles," *Journal of Agricultural and Food Chemistry*, vol. 45, no. 3, pp. 837–843, Mar. 1997, Publisher: American Chemical Society, ISSN: 0021-8561. DOI: 10.1021/jf9604807. [Online]. Available: https://doi.org/10.1021/jf9604807 (visited on 03/14/2024).

[106] J. Y.-L. Chan, S. M. H. Leow, K. T. Bea, W. K. Cheng, S. W. Phoong, Z.-W. Hong, and Y.-L. Chen, "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review," en, *Mathematics*, vol. 10, no. 8, p. 1283, Jan. 2022, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2227-7390. DOI: 10.3390/math10081283. [Online]. Available: https://www.mdpi.com/2227-7390/10/8/1283 (visited on 03/04/2024).

[107] P. Corcoran, "Optimal Configuration of Sensor Arrays for Odour Classification," en, *Measurement and Control*, vol. 30, no. 2, pp. 43–46, Mar. 1997, Publisher: SAGE Publications Ltd, ISSN: 0020-2940. DOI: 10.1177/002029409703000203. [Online]. Available: https://doi.org/10.1177/002029409703000203 (visited on 03/06/2024).

[108] A. N. Chaudry, T. M. Hawkins, and P. J. Travers, "A method for selecting an optimum sensor array," *Sensors and Actuators B: Chemical*, Proceedings of the International Symposium on Electronic Noses, vol. 69, no. 3, pp. 236–242, Oct. 2000, ISSN: 0925-4005. DOI: 10.1016/S0925-4005(00)00498-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400500004986 (visited on 03/04/2024).

[109] S. Zhang, C. Xie, D. Zeng, H. Li, Y. Liu, and S. Cai, "A sensor array optimization method for electronic noses with sub-arrays," *Sensors and Actuators B: Chemical*, vol. 142, no. 1, pp. 243–252, Oct. 2009, ISSN: 0925-4005. DOI: 10.1016/j.snb.2009.08.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400509006339 (visited on 03/04/2024).

[110] G. Wei, J. Zhao, Z. Yu, Y. Feng, G. Li, and X. Sun, “An Effective Gas Sensor Array Optimization Method Based on Random Forest,” in *2018 IEEE SENSORS*, ISSN: 2168-9229, Oct. 2018, pp. 1–4. DOI: 10.1109/ICSENS.2018.8589580. [Online]. Available: https://ieeexplore.ieee.org/document/8589580 (visited on 03/04/2024).

[111] J. Jung and J. D. Gibson, “The Interpretation of Spectral Entropy Based Upon Rate Distortion Functions,” in *2006 IEEE International Symposium on Information Theory*, ISSN: 2157-8117, Jul. 2006, pp. 277–281. DOI: 10.1109/ISIT.2006.261849. [Online]. Available: https://ieeexplore.ieee.org/document/4035966 (visited on 03/09/2024).

[112] J. Thorson, A. Collier-Oxandale, and M. Hannigan, “Using A Low-Cost Sensor Array and Machine Learning Techniques to Detect Complex Pollutant Mixtures and Identify Likely Sources,” *Sensors (Basel, Switzerland)*, vol. 19, no. 17, p. 3723, Aug. 2019, ISSN: 1424-8220. DOI: 10.3390/s19173723. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6749282/ (visited on 03/15/2024).

[113] M. Aliramezani, C. R. Koch, M. Secanell, R. E. Hayes, and R. Patrick, “An electrochemical model of an amperometric NOx sensor,” *Sensors and Actuators B: Chemical*, vol. 290, pp. 302–311, Jul. 2019, ISSN: 0925-4005. DOI: 10.1016/j.snb.2019.03.135. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400519305015 (visited on 03/15/2024).

[114] S. Li, M. Zhang, and H. Wang, “Simulation of gas sensing mechanism of porous metal oxide semiconductor sensor based on finite element analysis,” en, *Scientific Reports*, vol. 11, no. 1, p. 17 158, Aug. 2021, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-021-96591-2. [Online]. Available: https://www.nature.com/articles/s41598-021-96591-2 (visited on 03/15/2024).

[115] A. H. Goldstein and I. E. Galbally, “Known and Unexplored Organic Constituents in the Earth’s Atmosphere,” *Environmental Science & Technology*, vol. 41, no. 5, pp. 1514–1521, Mar. 2007, Publisher: American Chemical Society, ISSN: 0013-936X. DOI: 10.1021/es072476p. [Online]. Available: https://doi.org/10.1021/es072476p (visited on 04/10/2024).