DISSERTATION

A VERSATILE LOW-COST PLATFORM FOR PARTICULATE MATTER, VOLATILE ORGANIC COMPOUND, AND NOISE MONITORING

Submitted by

Emilio Molina Rueda

Department of Mechanical Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: John Volckens
Co-Advisor: Ellison Carter

Christian L'Orange
Shantanu Jathar
Ander Wilson

ABSTRACT

A VERSATILE LOW-COST PLATFORM FOR PARTICULATE MATTER, VOLATILE ORGANIC COMPOUND, AND NOISE MONITORING

Millions of people die prematurely each year from exposure to air pollution and other environmental hazards, and many more experience chronic disease. Mortality, morbidity, and other key indicators of adverse health are typically analyzed for large populations using city-, county-, or census track-level exposure data. A more local, or ideally, an individual-level understanding of human exposures would reduce uncertainty significantly and provide more actionable data. With the advent of affordable microelectronics, low-cost instruments for air quality and environmental monitoring are becoming increasingly common and more accessible to the public, but the data-quality gap between reference-grade monitors and low-cost alternatives is substantial. Combining the data-quality standards of reference-grade monitors with the accessibility of new consumer electronics would advance the state-of-the-art in environmental monitoring. This work describes the development and validation of a platform to monitor and collect particle and gas pollutants, and other environmental factors, at a lower cost than established technologies, while preserving the rigor and data-quality objectives associated with reference-grade methods. We developed a platform based on the AirPen, a low-cost monitor for assessing personal exposure to particulate matter and volatile organic compounds. Variations of the platform with different form factors, sensors, and sampling media compatibility were designed, and then tested in the field. Our platform implemented previously validated methods such as NIOSH method 0500 for total particulate matter and EPA method TO-11A for carbonyls.

The sampling throughputs achieved in two field studies were considerably higher (5x – 10x) than those typically achieved with conventional instruments within similar timeframes and populations. The spatial and temporal coverage of our results revealed insights on exposure that could be missed by more sparse monitoring. Results from this work demonstrate that new technology can bridge the data-quality gap and address the barriers of entry (e.g., cost, ease of use), to complement the limited regulatory (reference-grade) monitoring efforts currently in place.

ACKNOWLEDGEMENTS

My achievements have been possible by virtue of my parents' boundless support. Their hard work allowed me to travel light through life and chase the most enticing goals. My mom Ivannova taught me the importance of integrity and set an example of excellence that I try to meet. My sister Estefi has always backed me enthusiastically and been there for me when I needed it most. My grandparents were a key part of my upbringing, so my success is theirs. My family's unconditional love has helped me face the ups and downs of graduate school without faltering. All of them have endured five years of being four thousand miles away (as I have been reminded many times), yet they always received me with a big hug and a smile on each of my visits. Thank you for your patience and understanding.

My advisors John Volckens and Ellison Carter have been essential in my academic journey. I met Ellison for the first time in 2018, when she organized a workshop to bring researchers from around the world together. During the event, many early career researchers like me had the opportunity to learn, share experiences, and visit Colorado for the first time. I fell in love with Colorado during that trip. I met John in early 2020 when CSU's mechanical engineering department invited me to visit them. John introduced me to members of his team, and I had the chance to talk to them at length. My impressions were of warmth, respect, and transparency. This was very reassuring, especially when I had to make a life-changing decision to move far from home during a time of global chaos. During my years at CSU, John and Ellison have always been supportive and fair, guiding me when I was lost and giving me freedom to learn in the ways that fit me best. They are committed to providing their teams with state-of-the-art resources and a healthy working environment. I will be forever grateful.

My committee members, Christian L'Orange, Shantanu Jathar, and Ander Wilson have contributed to my academic formation with solid theoretical foundations and vast empirical experience. While working with them in a variety of projects, I have acquired valuable skills that make me a better researcher. I would like to acknowledge Christian, in particular, for going above and beyond to help me with all sorts of matters, from the menial paperwork that robs us valuable research time, to complex experimental methods, and the prototyping skills which are the backbone of our work. I am a better engineer and scientist after working with Christian.

I would also like to recognize the impact that other mentors have had on my development. First and foremost, Jessica Tryner is not only responsible for a large portion of the development of the technology covered in this dissertation, but also the person who worked closest to me for several years and shared a vast amount of invaluable knowledge. Jess is one of the best mechanical engineers I have ever met, an outstanding writer, and a data analysis genius. More importantly, she has huge sense of justice and empathy, she is a great friend, and a multitalented adventurer. Being mentored by Jess has been a highlight of my time in Colorado. Alfredo Valarezo and Lorena Bejarano have also been my mentors since I was an undergraduate student at USFQ many years ago. As professors, they contributed to my initial engineering training. Then, I wrote my undergraduate thesis under Alfredo's advising, and I joined Alfredo's and Lore's research team shortly after. They always trusted in me and the opportunities they offered me paved the way for all the things that came later. My time at USFQ was the happiest of my life.

My peers, coworkers, and research partners have also been a vital part of this work. My fellow students, Ky Tanner, Julian Probsdorfer, and Jacob Fontenot were always willing to give me a hand when I was overwhelmed. Jane Andales worked tirelessly along me to achieve seemingly impossible deadlines. Jane's and Amy Sullivan's analytical frameworks are indispensable for the

platform we developed. Casey Quinn wrote the original firmware for the AirPen and contributed to subsequent versions. Meghan Cosgrove, Elizabeth Williams, Ashley Anderson, and Marilee Long contributed to the AirPen occupational exposure research. The Environmental Defense Fund partnered with us for the ambient monitoring research.

None of this would have been possible without the support of many volunteers and participants that collected data and samples for testing, validation, and field demonstrations. Researchers and staff at the Western Colorado Research Center were unbeatable participants, wearing the first version of the AirPen and traditional samplers for a week; a significant burden that I hope will not be needed again. The participants of our full-scale occupational exposure deployment followed the wearing protocols for a full work-shift admirably well. The hosts of our Houston ambient monitoring deployment collected samples and shipped them weekly for nearly two months and were always eager to learn and contribute to the project.

Last but not least, I thank my friends for keeping me sane during this challenging era of my life. Katie and Nathan, you were the most unexpected gift that Fort Collins could give me. We are somehow so different and so similar, and we met by pure luck. It is fun to be around you regardless of the plan. Thank you for your kindness and hospitality. Jess (again), Tim, Jeff, and Christine took me to breathtaking places in Colorado where most locals have never been. You are the only ones who match (and sometimes exceed) my adventurer spirit, and you are outstanding scientists and human beings. Juli found me (the only other USFQ dragon in town?) with her forensic skills and kept me around with Ecuadorian food, fun activities, and good vibes. Dani and Andre have been the best neighbors I could ever ask for. You always manage to take me out of my cave to decompress (even when I really should be working). And Estefy, thank you for not becoming a stranger and having food with me every time I go home.

DEDICATION

Dedicated to the teachers, professors, and mentors who, instead of telling me what to think, went above and beyond to show me how to do it. In particular, to Alfredo Valarezo and Lorena Bejarano, who saw potential behind my divergent nature and motivated me to pursue my academic curiosity. Thank you for offering me a place on your team. You changed my life for the better.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Breathing air is the most basic human need. The correct blend of innocuous nitrogen and life-supporting oxygen allows us to think, work, exercise, digest, and sleep well. But, along with nitrogen, oxygen, and some inert gases, we breathe small amounts of harmful gases and airborne particles that have major effects on our health and overall quality of life. According to the Institute for Health Metrics and Evaluation (IHME), ~8 million deaths per year across the globe were attributable to exposure to air pollution for all years from 1998 to 2021.[1] Particulate matter is the main contributor to air pollution-related mortality and morbidity,[2–4] warranting the most oversight of any air pollutant. However, biogenic and anthropogenic emissions, as well as chemical reactions in the atmosphere, create complex mixtures of hundreds of air contaminants, many of which are rarely studied. Furthermore, exposures to air pollution and other hazards vary widely between and within populations.[5–11] Traditional characterization of exposure at city-level or by broad population groups (e.g. by occupation) leads to uncertainty in the analysis of health effects.

In the era of genomics, metabolomics, and many other *omics*, a comprehensive understanding of exposures to numerous air pollutants and environmental factors at an individual level (i.e., exposomics) still seems unattainable. Nevertheless, community, household, and personal multi-exposure monitoring are becoming increasingly relevant as technology enables the development of low-cost, miniaturized instruments. With new devices becoming available rapidly and promising ever-growing functionality, keeping track of all of them and evaluating their true performance is an emerging challenge. Unfortunately, the accuracy and reliability of reference-grade monitors is still uncontested for most applications.[12,13] Low-cost sensors for PM and gases

are typically prone to bias, drift, and poor selectivity, and require complex corrections to make their outputs moderately accurate.[14–18]

To bridge the performance gap between traditional (i.e., reference-grade) instruments and their low-cost counterparts, researchers and manufacturers must apply the same rigor that led to the establishment of the former, when developing the latter. Scientific and technical rigor, thorough validation of the technology, in addition to continuous quality control and assurance, are indispensable. Moreover, the burden of operating monitors and validating data should be manageable for non-experts or automated as much as possible to facilitate wide adoption.

If reliability and barriers of entry (i.e., cost, complexity of operation) are addressed, the coverage and data quality needed to revolutionize our understanding of exposure and its health effects could be within reach. The AirPen was designed to tackle various of these challenges by combining (in a wearable form-factor) time-integrated sampling technology based on reference methods (e.g., gravimetry for PM, thermal desorption and gas chromatography for VOCs) and a suite of sensors for time-resolved monitoring, automated control, and data/sample quality assurance.[19] As analytical frameworks were developed and instrument and sampling media variations were introduced, the AirPen became a platform for a range of sampling applications.

This dissertation covers three aspects of the research and development process, as well as the adaptations and evolution of the AirPen as a versatile platform for multi-hazard monitoring. Chapter 2 presents the findings of a comprehensive field evaluation of low-cost light scattering particulate matter sensors that informed the component selection for the AirPen and showed the strengths and fundamental limitations of light scattering sensors for PM monitoring. Chapter 3 covers the learnings of a facility-wide deployment of the AirPen at a furniture manufacturing facility as a wearable sampler for assessing occupational exposure to dust, noise, and

formaldehyde. That work demonstrates a framework for the sensor-based validation of sampling results, in addition to highlighting the importance, through statistical analyses, of larger sample sizes in occupational exposure assessment. Finally, chapter 4 describes the development of a stationary, battery-powered version of the AirPen for hazardous air pollutant monitoring and the results of a community-based deployment during hurricane season in Houston, TX.

CHAPTER 2. SIZE-RESOLVED FIELD PERFORMANCE OF LOW-COST SENSORS FOR PARTICULATE MATTER AIR POLLUTION[1]

Chapter overview

Particulate matter (PM) air pollution is a major health hazard. The health effects of PM are closely linked to particle size, which governs its deposition in (and penetration through) the respiratory tract. In recent years, low-cost sensors that report particle concentrations for multiple-sized fractions ($PM_{1.0}$, $PM_{2.5}$, PM_{10}) have proliferated in everyday use and scientific research. However, knowledge of how well these sensors perform across the full range of reported particle size fractions is limited. Unfortunately, erroneous particle size data can lead to spurious conclusions about exposure, misguided interventions, and ineffectual policy decisions. We assessed the linearity, bias, and precision of three low-cost sensor models, as a function of PM size fraction, in an urban setting. Contrary to manufacturers' claims, sensors are only accurate for the smallest size fraction ($PM_{1.0}$). The $PM_{1.0–2.5}$ and $PM_{2.5–10}$ size fractions had large bias, noise, and uncertainty. These results demonstrate that low-cost aerosol sensors (1) cannot discriminate particle size accurately and (2) only report linear and precise measures of aerosol concentration in the accumulation mode size range (i.e., between 0.1 and 1 μm). We recommend that crowdsourced air quality monitoring networks stop reporting coarse ($PM_{2.5–10}$) mode and PM_{10} mass concentrations from these sensors.

[1] Reprinted (or 'Reprinted in part') with permission from "Size-Resolved Field Performance of Low-Cost Sensors for Particulate Matter Air Pollution". Emilio Molina Rueda, Ellison Carter, Christian L'Orange, Casey Quinn, and John Volckens; Environmental Science & Technology Letters 2023, 10, 3, 247-253; DOI: https://doi.org/10.1021/acs.estlett.3c00030; Copyright 2023 The Authors.

Introduction

Airborne particulate matter (PM) is a major public health concern. Exposure to PM contributes to over a million premature deaths worldwide, and populations subject to long-term exposure suffer from significantly higher cardiovascular and respiratory morbidity.[2,20,21] The atmospheric fate and transport of PM is largely determined by particle size (d_p, in μm), as is the penetration and deposition of PM within the human respiratory tract.[22] How particle size relates to PM health effects remains an active area of research; human exposure to size modes such as ultrafine (d_p < 0.1 μm), fine (d_p < 2.5 μm), and coarse (2.5 < d_p < 10 μm) PM have each been associated with adverse health outcomes.[23,24] Traditional (reference) methods for measuring size-resolved particle concentrations are expensive (e.g., instrument costs from $10,000 to $100,000 each) and resource intensive,[25] limiting their use to spatially-sparse outdoor monitoring networks in high-income countries and research studies with short sampling durations and small samples sizes.

The emergence of low-cost PM sensors (miniaturized, mass-produced devices that cost ~$15 to $50 each) has facilitated the deployment of cheaper (~$250/each) monitors in crowdsourced measurement networks (e.g., PurpleAir, Clarity) that are denser than traditional national/regional-scale networks.[26] These networks are being leveraged to support growing interest in PM exposure and health science globally. Most low-cost PM sensors operate on the principle of aerosol light scattering: a fan draws PM into a small housing, the PM passes through a focused beam of light, and a photodetector measures the intensity of the light scattered by the particles.[26] Most low-cost PM sensors report mass and number concentrations across a range of sizes (e.g., $PM_{0.3\text{-}0.5}$ $PM_{0.5\text{-}1.0}$, $PM_{1.0\text{-}2.5}$ $PM_{2.5}$, PM_{10}). By including concentrations across multiple size bins as sensor outputs, manufacturers specify, either explicitly or implicitly, that their sensors can classify particles by size. However, manufacturer documentation often omits the

working principle of the sensor (i.e., whether it functions as an optical particle counter or a nephelometer) and the bias/precision of size-resolved outputs.

Particle sizing is challenging even for reference-grade instruments, especially those that rely only on light scattering.[27] For example, the signal produced by a given particle can be below the limit of detection, saturated, ambiguous (e.g., when two particles coincide in the detector), or inherently biased by aspiration or transmission errors during sampling. Sensor evaluations carried out by manufacturers and third parties generally involve evaluating the accuracy and precision of *cumulative* mass concentrations (e.g., $PM_{1.0}$, $PM_{2.5}$, PM_{10}). This approach does not provide sufficient information to assess a sensor's ability to quantify and classify particles in specific size ranges. Laboratory experiments and physics-based models have shown that most low-cost PM sensors detect particles larger than 1 μm with low efficiency.[28–30] Additionally, field-based data suggest that low-cost PM sensors underestimate ambient concentrations of wind-blown dust (which likely includes many particles > 1 μm).[14,31]

Here, we compare differential mass concentrations (e.g., $PM_{1.0\rightarrow2.5}$) from three low-cost PM sensors to differential mass concentrations measured by a federal equivalent method (FEM) PM monitor in an outdoor urban environment to better understand the ability of low-cost sensors to detect PM in different size fractions. We focus on the most common PM size ranges (<1.0, 1.0-2.5, and 2.5-10 μm) and compare performance metrics for differential mass to cumulative mass to show how the latter can mask sensor limitations.

Materials and methods

Low-cost sensors

We evaluated three light-scattering sensor models (two units of each model) with prices and form factors that would be appropriate for use in large networks and/or personal exposure monitors (i.e., sensors that cost < $100 and weigh < 50 g): the PMS5003 (Plantower, Beijing, China), the SPS30 (Sensirion, Stäfa, Switzerland), and the IPS-7100 (Piera Systems, Mississauga, Canada). Specifications for all sensors are summarized in Table 1, and additional descriptions and data logging setups are available in Appendix A (see sections 1 and 2). All sensors were new and operated according to manufacturer recommendations (without additional calibration).

Table 1: Low-cost sensors specifications.

	Plantower PMS5003	**Sensirion SPS30**	**Piera IPS-7100**
Stated Outputs			
Mass concentration	$PM_{1.0}$, $PM_{2.5}$, PM_{10}	$PM_{1.0}$, $PM_{2.5}$, PM_4, PM_{10}	$PM_{0.1}$, $PM_{0.3}$, $PM_{0.5}$, $PM_{1.0}$, $PM_{2.5}$, PM_5, PM_{10}
Number concentration	$PC_{0.3}$, $PC_{0.5}$, $PC_{1.0}$, $PC_{2.5}$, PC_5, PC_{10}	$PC_{0.5}$, $PC_{1.0}$, $PC_{2.5}$, PC_4, PC_{10}	$PC_{0.1}$, $PC_{0.3}$, $PC_{0.5}$, $PC_{1.0}$, $PC_{2.5}$, PC_5, PC_{10}
Cost at quantity of 1	$15	$50	$80
Weight	42 grams	26 grams	26 grams
Size	50 x 38 x 21 mm	41 x 41 x 12 mm	46 x 42 x 12 mm

Reference monitors

A GRIMM EDM 180 (GRIMM Aerosol Technik, Ainring, Germany) was chosen as the reference monitor due to its ability to measure PM in 31 particle size channels (0.25 – 32 μm). The EDM 180 is an approved federal equivalent method (FEM) monitor for $PM_{2.5}$ when operated continuously at a 1.2 L min^{-1} volumetric flow rate with a Nafion air dryer inside an isothermal inlet.[32,33] We complied with this configuration and added the fully temperature-controlled

weather protection housing suggested by the manufacturer. This monitor is also UK-MCERTS certified for PM_{10}.[34]

For quality assurance, we compared $PM_{2.5}$ and PM_{10} measurements between the GRIMM EDM 180 and a collocated Beta Attenuation Monitor (5014i, Thermo Scientific, Waltham, MA, USA) (Figure A5). Additionally, we obtained ambient temperature, humidity, barometric pressure, and wind speed data from a collocated weather station (Vantage Pro2, Davis Instruments, Hayward, CA, USA) (see weather data in Appendix A section 7).

Field deployment

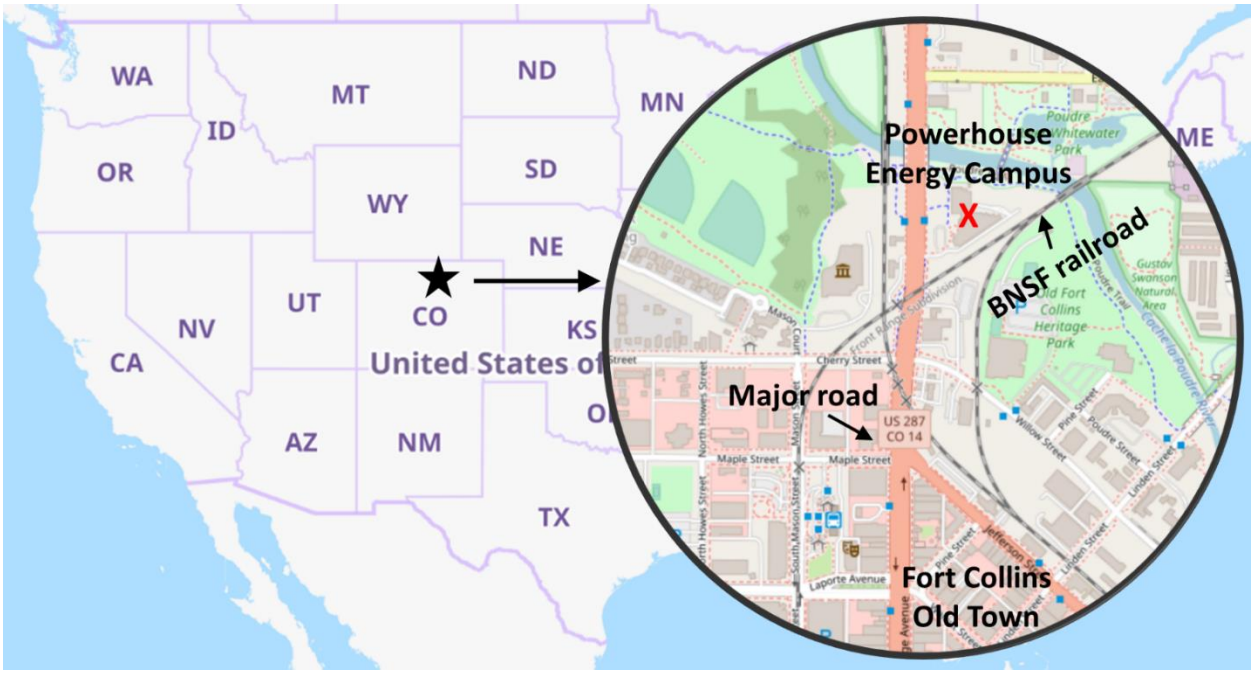

Figure 1: Map of the Powerhouse Energy Campus building in Fort Collins, where the sensors were tested.

The instruments were deployed on the roof of the Colorado State University Powerhouse Energy Campus in Fort Collins, Colorado, USA (Figure 1). The building is located in an urban environment, adjacent to a major road and a railway, and experiences seasonal dust and wildfire events, despite having a relatively low background PM concentration. The GRIMM EDM 180

ran continuously on a roof above the second floor of the building. The low-cost sensors were installed on one side of the reference monitor (Figure 2).

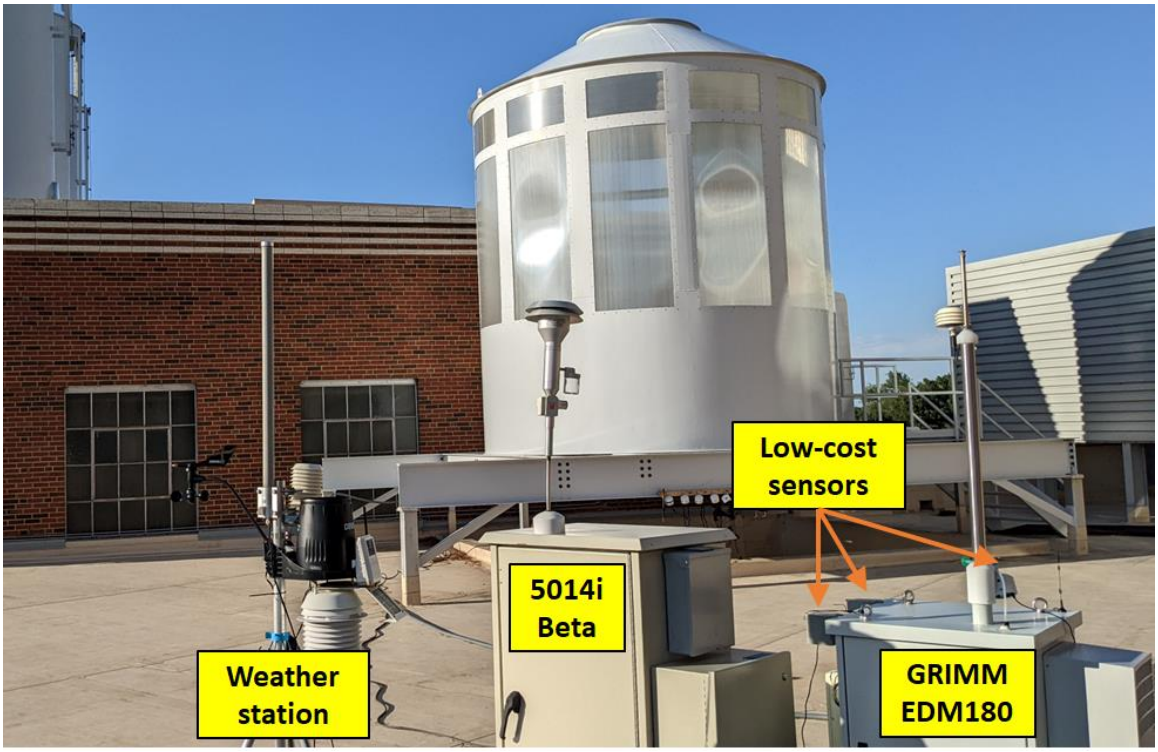

Figure 2: Layout of the instruments used for the field evaluation.

Data processing

Data were collected during two periods with different weather: a fall/winter dataset that spanned from November 23, 2021 to January 9, 2022 (48 days) and a summer dataset that spanned from June 13, 2022 to July 30, 2022 (48 days). Data were aggregated to 1-hour averages and PM mass concentrations were calculated for each monitor in the following size ranges: $PM_{1.0}$, $PM_{2.5}$, PM_{10}, $PM_{1.0 \rightarrow 2.5}$, and $PM_{2.5 \rightarrow 10}$ (see variable definitions in Appendix A section 3).

Statistical analyses

Descriptive statistics, performance metrics, regression models, and graphical tools (e.g., scatter plots) were used to assess sensor precision (coefficient of variation among co-located devices of

the same model), linearity (coefficient of determination vs. reference), and bias (e.g., RMSE, MAE, NMB vs. reference) as a function of particle size range. A complete list and details of the performance metrics and statistical methods used in this study are available in Appendix A (sections 4 and 5).

Results and discussion

Table 2: Size-resolved descriptive statistics and performance metrics over the full field evaluation (fall/winter and summer periods combined). Parameters marked as N/A correspond to models that violated linear regression assumptions.

	$PM_{1.0}$	$PM_{2.5}$	PM_{10}	$PM_{1.0-2.5}$	$PM_{2.5-10}$
Descriptive statistics (EDM180)					
Median [µg/m³]	2.9	4.9	12.3	1.4	6.4
Mean [µg/m³]	4.4	6.5	17.8	2.0	11.4
IQR [µg/m³]	4.0	5.4	16.4	1.7	13.1
Range [µg/m³]	0 – 119.9	0.1 – 124.7	0.1 – 351.4	0 – 51.4	0 – 295.4
Coefficient of determination (R^2)					
Piera IPS-7100	0.83	0.65	0.07	0.01	0.07
Sensirion SPS30	0.93	0.72	0.23	0.09	0.22
Plantower PMS5003	0.90	0.76	0.07	0.01	0.00
Slope | Intercept					
Piera IPS-7100	0.87 | -1.54	1.70 | -6.51	N/A	N/A	N/A
Sensirion SPS30	0.88 | -0.54	0.78 | -1.39	N/A	N/A	N/A
Plantower PMS5003	1.44 | -1.01	1.68 | -3.01	N/A	N/A	N/A
Mean absolute error [µg/m³]					
Piera IPS-7100	2.36	3.87	13.89	2.34	10.42
Sensirion SPS30	1.20	3.00	14.10	1.84	11.26
Plantower PMS5003	1.68	3.45	12.88	2.07	10.88
Normalized mean bias					
Piera IPS-7100	-47.8 %	-31.0 %	-65.7 %	5.2 %	-85.3 %
Sensirion SPS30	-24.4 %	-43.7 %	-78.9 %	-85.3 %	-98.8 %
Plantower PMS5003	21.0 %	21.6 %	-47.2 %	22.9 %	-86.2 %
Coefficient of variation					
Piera IPS-7100 (2 units)	16.0 %	20.4 %	18.2 %	28.0 %	17.3 %
Sensirion SPS30 (2 units)	6.0 %	6.1 %	6.3 %	8.0 %	50.9 %
Plantower PMS5003 (2 units)	22.2 %	16.9 %	13.4 %	16.3 %	22.2 %

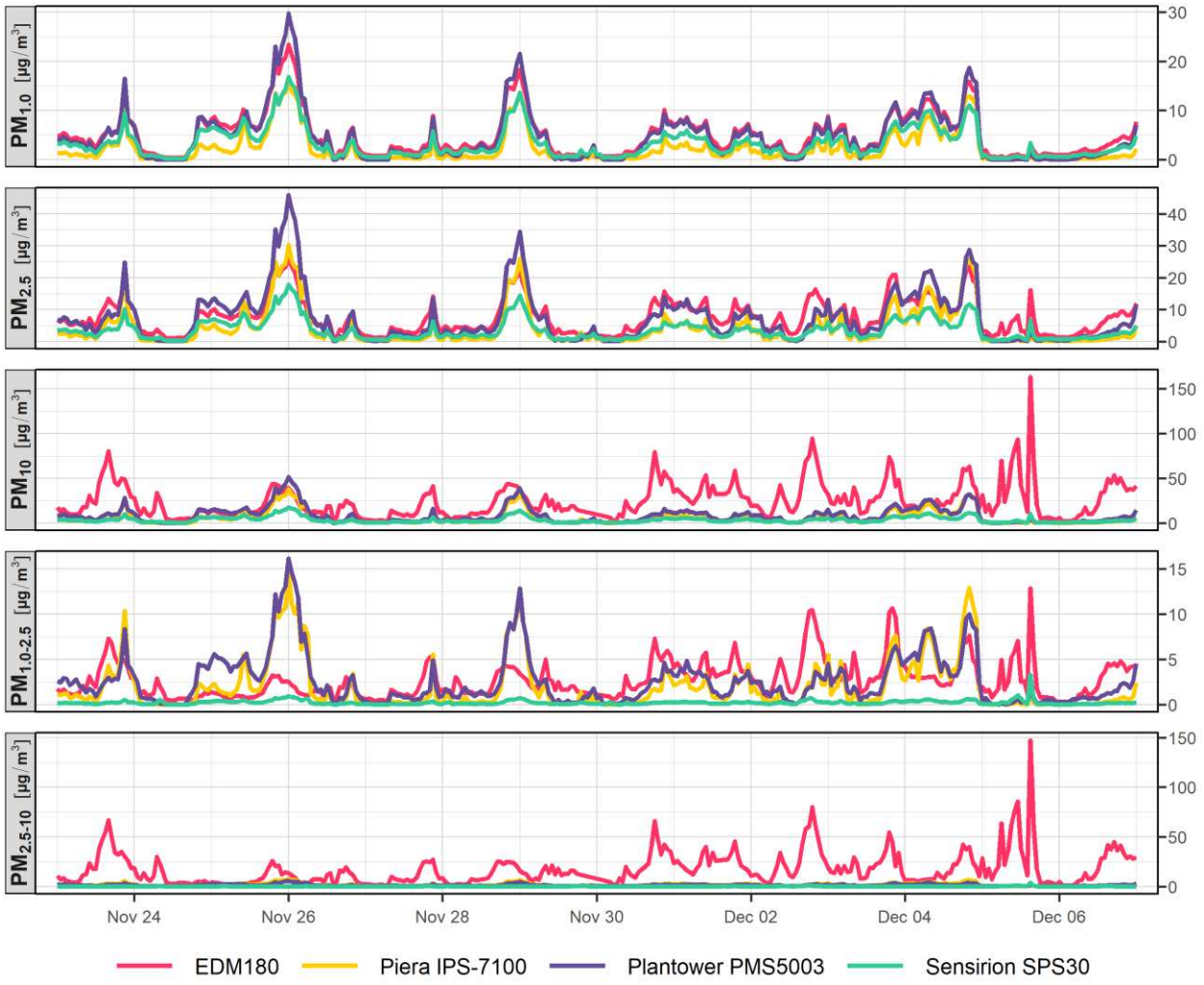

Figure 3: Time series graph of PM concentrations (cumulative and differential). A subset composed of the first 14 days of the fall/winter period is shown to improve readability. A similar plot for the summer period is available in Appendix A (Figure A6).

Marked differences were evident in the low-cost sensors' responses to PM of different size fractions (Figures 3 and 4). The low-cost sensors measured $PM_{1.0}$ with relatively low bias (mean absolute error [MAE] ranging from 1.2 to 2.4 µg/m^3; Table 1) and strong linear correlation (R^2 from 0.83 to 0.93). For $PM_{2.5}$, the sensors showed slightly worse agreement with the reference monitor; MAE increased by a factor of two for all sensors (3.0 to 3.9 µg/m^3) and linear correlation decreased ($0.65 \leq R^2 \leq 0.76$ compared to $0.83 \leq R^2 \leq 0.93$, as stated above). A larger disparity was evident for PM_{10} concentrations reported between the low-cost sensors and

reference monitor (R^2 from 0.07 to 0.23). Consistently low PM_{10} concentrations reported by the low-cost sensors, relative to the reference monitor, indicated that larger particles (i.e., $d_p > 1.0$ μm) were not adequately detected by the sensor models we evaluated. Differential mass concentrations provided further evidence of this limitation. The $PM_{1.0 \rightarrow 2.5}$ and $PM_{2.5 \rightarrow 10}$ signals were noisy and nearly uncorrelated with the reference measurement throughout the experiment. Some of these patterns (or lack thereof) were clearly identifiable from a simple visual assessment of the time series plots (Figure 3).

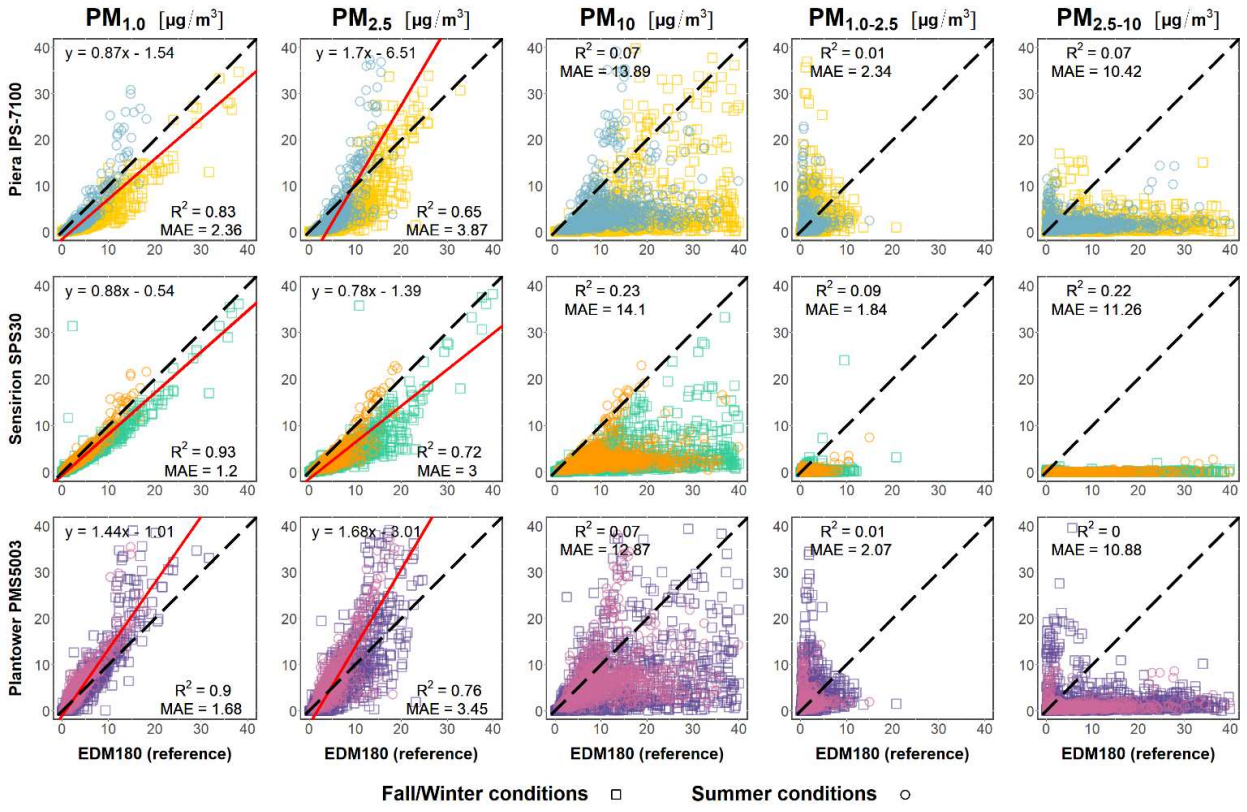

Figure 4: Regression plots of low-cost sensor PM estimates versus the EDM180 reference measurements. Dashed black lines are identity (1:1) lines and continuous red lines are the lines of best fit (for models that complied with linear regression assumptions). All the hourly averages (n = 1763) were used to compute the regression models, but some points are not shown in the plots due to the axes range.

The readings from all sensors lost accuracy when going from PM_1 to $PM_{2.5}$, and more so when going from $PM_{2.5}$ to PM_{10}. In those same plots, the $PM_{2.5 \rightarrow 10}$ signals from the low-cost sensors

remained flat and nearly zero throughout the campaigns, despite the reference monitor reporting $PM_{2.5\rightarrow10}$ concentrations of ~10 – 50 μg/m^3. This unresponsiveness is consistent with a fundamental inability to sense and/or classify particles in that size range (2.5 – 10 μm), which has been demonstrated previously for some of these same low-cost sensor models in laboratory experiments[29] and physical optical models.[30] The regression plots in Figure 4 and performance metrics shown in Table 2 confirm that all the low-cost sensors tested here miss almost all particle mass in the $PM_{2.5\rightarrow10}$ size range under the real-world, outdoor conditions in which these sensors are often used.

Additionally, every sensor demonstrated poor detection of the intermediate differential size fraction (1.0 – 2.5 μm). As shown in Figure 4, $PM_{1.0\rightarrow2.5}$ errors appear random with no clear trends or consistent bias. Qualitatively, this was seen as large scattering of data points in the regression plots (Figure 4) and the REU plots (Figure A2). Quantitatively, the very low coefficients of determination ($R^2 < 0.1$ for all linear regression models) and large errors were coherent with a signal composed mostly of noise. Our results indicate that these sensors can estimate ambient PM mass concentrations in the accumulation mode ($0.1 < d_p < 1$ μm), but that they cannot reliably detect particle mass in the 1.0 – 2.5 μm or the 2.5 – 10 μm size ranges. Our results are consistent with Ouimette et al.,[30] who developed a physical-optical model of the PMS5003 and theorized that this device would struggle to size-classify particles due to signal misclassification from multi-angle scattering within the instrument's detection zone. Laboratory evaluations of the SPS30 and the PMS5003 have found that size detection ranges don't adhere to manufacturer specifications and that size distribution data from these sensors are unreliable.[28,29] Our results under real-world conditions confirm these reports. The combined evidence suggests that the size-resolved data reported by these sensors is based on mathematical artifices rather

than on real measurements of size distribution. There are several reasons why these optically based, low-cost sensors fail to respond adequately to particles larger than 1 μm. For instance, the sensing zones in these instruments have truncated viewing angles that fail to capture forward light scattering from particles.[30] Such *truncation error* has been shown to produce a dramatic loss in signal as particle size increases from 0.5 to 5 μm. Furthermore, these devices are likely to experience inertial losses during aspiration and transmission of particles from ambient air to their respective sensing zones.[35]

Ideally, the sensors should provide linear response with relatively low noise (i.e., good precision). Non-zero intercepts and non-unity slopes in the regression calibration models are not a major concern because optical instruments are sensitive to aerosol characteristics such as shape and refractive index, which means that they need to be calibrated to specific sampling conditions in applications where high accuracy is needed.[36] For $PM_{1.0}$, most points on the regression plots are close to the 1:1 line indicating good agreement with the reference monitor, and all the sensors fit the linear model reasonably well ($R^2 \geq 0.83$; see also Figures A7, A8). Additionally, the relatively small scatter in the $PM_{1.0}$ regression plots and the narrow band patterns in the REU plots (Figure A2) denote low noise. The $PM_{2.5}$ estimates combine the good performance of $PM_{1.0}$ with the random noise of $PM_{1.0\rightarrow2.5}$. This translates to more scattering, larger errors, and poorer fit of the linear models ($0.65 \leq R^2 \leq 0.76$) but, depending on the application, a calibrated sensor could produce usable $PM_{2.5}$ estimates, as has been reported previously.[37–39] For instance, to evaluate $PM_{2.5}$ data quality, the uncertainty could be analyzed as the European Commission recommends, which specifies a 50% REU upper bound as the data quality objective (DQO) for low-cost sensors.[40] Alternatively, $PM_{2.5}$ could be estimated from the more accurate $PM_{1.0}$ output by developing *ad-hoc* conversion equations for a combination of sensor model and use-case.

On the other hand, the PM_{10} signals of these sensors are clearly flawed, incorporating the noise of $PM_{1.0 \rightarrow 2.5}$ and the systematic bias of $PM_{2.5 \rightarrow 10}$. The regression plots of PM_{10} show high scattering, large bias, and no linear or other type of trend. The REU plots of PM_{10} show extremely large uncertainty with a combination of bias and noise across all concentrations (Figure A2). Consequently, the PM_{10} signal of the sensors we tested should be disregarded, as it appears to provide no meaningful output. A field evaluation and calibration by Kosmopoulos et al. reached similar conclusions and recommended the removal of high PM_{coarse} events from datasets collected with the PurpleAir to improve the accuracy of the PM_1 and $PM_{2.5}$ signals, but the authors didn't find a way to obtain adequate PM_{10} estimates.[41]

Some studies where only cumulative concentrations (e.g., $PM_{1.0}$, $PM_{2.5}$) were analyzed have found good agreement between low-cost sensors and FEM monitors for $PM_{2.5}$ and even for PM_{10}.[42,43] Nevertheless, our analysis of the differential concentrations (i.e., $PM_{1.0\text{-}2.5}$ and $PM_{2.5\text{-}10}$) shows that the $PM_{2.5}$ performance is driven largely by the performance in the $PM_{1.0}$ range. If sensors are tested in environments where $PM_{1.0}$ constitutes a substantial proportion of $PM_{2.5}$ and PM_{10}, or where the particle size distributions resemble the conditions that the manufacturer used to calibrate the devices, the accuracy of $PM_{2.5}$ and PM_{10} signals can be artificially high. Hence, evaluations do not provide a full picture of sensor performance if a wide range of conditions is not covered, and if only cumulative concentrations are analyzed.

Our work does not focus on calibration schemes for improving low-cost sensor data, for which there are many potential strategies.[31,36,37,44,45] Further, we tested only three different sensor models; however, these models typify the state-of-the-art for low-cost light scattering devices and represent the majority of in-use technologies for crowdsourced measurement networks

around the world.[46–48] Although our results are limited to a single geographic location and two seasons, they are consistent with previous theoretical and laboratory investigations.[27,29,30]

The limitations of low-cost sensors presented here should be acknowledged by sensor manufacturers (or product integrators) and the networks that leverage these sensors should cease reporting PM_{10} mass concentrations. Use of inaccurate particle size distribution data for research applications such as source apportionment or outdoor-to-indoor air penetration[49] could lead to wrong conclusions and, ultimately, to misguided public health interventions. For everyday applications where people want to understand local PM_1, $PM_{2.5}$, and PM_{10} sources and levels, it's important to understand how effective (or ineffective) low-cost sensors are at detecting particulate matter depending on the primary sources (e.g., wildfire smoke, vehicle exhaust, cooking aerosol, wind-blown dust) and sizes of the particles. In all cases, clear guidelines for using (or disregarding) sensor outputs based on their trustworthiness (i.e., accuracy, precision, uncertainty) will benefit the user community. Specific applications where low-cost sensors have been shown to produce reliable estimates of aerosol mass concentration include calibration schemes for which the following conditions hold: (1) the aerosol of interest is stable in terms of size and refractive index, (2) the environmental conditions are accurately recorded (especially relative humidity), and (3) the aerosol mass median diameter (MMD) falls within the accumulation mode (i.e., $0.1 < MMD < 1$ μm). Applications where these conditions have been met may include urban fine particulate matter,[50,51] wildfire smoke,[52,53] and household solid fuel burning.[54] When calibration schemes fail to account for changes in particle size, particle refractive index, and ambient relative humidity, the response from these sensors will be uncertain and subject to bias.[36,55] Whenever possible, calibration schemes should be designed to account for potential variability in particle characteristics and ambient conditions, as noted above.

Environmental conditions (i.e., temperature, relative humidity, barometric pressure) during calibration should ideally span ranges typical of the environmental conditions under which the sensors will be deployed in their intended setting(s). Sensor calibration should also occur in the presence of air pollution sources that are similar in composition and magnitude to those that the sensors will experience during deployment. Regardless of the calibration scheme, fundamental safeguards for the use of sensors should include protocols that address data handling and initial processing, outlier detection and removal, sensor detection limit, and data completeness. Further, calibration schemes should disclose the kind of reference air quality monitor used, the duration of the co-location experiment and ambient conditions, the time-averaging intervals used for processing the data, and the statistical model(s) selected with appropriate justification (e.g., evidence of model assumptions being met).[36,45]

In summary, we conclude that low-cost PM sensors commonly used in crowdsourced measurement networks are best described as *accumulation mode* PM ($PM_{0.1\text{-}1.0}$) sensors with little-to-no sizing ability or measurement reliability outside this size range. None of the devices evaluated here could detect coarse mode PM ($PM_{2.5\text{-}10}$) and most struggled to measure the 1.0 – 2.5 μm range when compared to a size-resolved reference monitor. Therefore, $PM_{2.5}$ estimates from these low-cost sensors should also be interpreted with caution and PM_{10} estimates from these sensors should be seen only as a proxy representing the contribution of the accumulation mode to the PM_{10} fraction.

CHAPTER 3. FACILITY-WIDE ASSESSMENT OF OCCUPATIONAL EXPOSURE TO DUST, NOISE, AND FORMALDEHYDE[2]

Chapter overview

Exposure to particulate matter, hazardous gases, and noise among workers causes millions of deaths, injuries, and disabilities annually. These exposures are not measured frequently, and most workers are never monitored because resources for monitoring are limited globally. Even in high-income countries, just two or three samples per workplace inspection are collected on average. Considerably larger sample sizes are required to estimate measures of central tendency and upper percentiles of common exposure distributions. Therefore, interventions and epidemiological studies often rely on poor estimates of those parameters. The objective of this study was to demonstrate that a small team can cost-effectively measure all workers' exposures to multiple hazards within a facility (~100 workers) in a single day while minimizing participant burden and loss of productivity. We deployed novel, compact personal monitors—called AirPens—to measure exposures to total particulates, formaldehyde, and A-weighted noise among workers at a furniture manufacturing facility during one full work shift. AirPens were preloaded with sampling media and preprogrammed to minimize field deployment time. On-board sensor data were used to identify sampling anomalies, confirm that AirPens were worn, and validate sampling results. Valid samples were then used to estimate exposure distribution parameters. A team of five people deployed 83 AirPens at the facility. After quality assurance screening, 67

[2] This work was submitted to the Journal of Exposure Science and Environmental Epidemiology on March 5, 2025 for peer-review as "Facility-wide assessment of occupational exposure to dust, noise, and formaldehyde". Emilio Molina Rueda, Ellison Carter, Christian L'Orange, Meghan Cosgrove, Elizabeth Williams, Ashley A. Anderson, Jane Andales, Jessica Tryner, Marilee Long, Ander Wilson, and John Volckens.

PM, 67 noise, and 22 formaldehyde samples remained valid (or 72, 67, and 31, respectively if samples below the limit of detection are counted as valid). PM and noise exposure distribution parameters (e.g., arithmetic and geometric mean, 95th percentile) were estimated with high certainty. Formaldehyde results were less reliable due to the smaller sample size. Use of streamlined multi-hazard monitoring technology enables dramatically higher sample throughput for workplace exposure assessment. This approach reduces the uncertainty of workplace risk assessment and facilitates more in-depth analyses of personal exposure.

Introduction

Occupational risks contribute substantially to global morbidity and mortality. The World Health Organization (WHO) and the International Labour Organization (ILO) estimate that nearly 1.9 million deaths were caused by work-related diseases and injuries in 2016, globally.[56,57] Over 40% of the work-related mortalities were linked to chronic respiratory diseases and various types of cancer, which, in turn, were partially attributable to exposure to occupational particulate matter (PM) and gases.[56,58] Workers in many industrial sectors also experience long-term exposure to high levels of noise (> 85 dBA), which increases the incidence of hearing loss by at least a factor of 2.[59,60] Furthermore, evidence suggests that noise exposure is associated with elevated risk of cardiovascular disease.[61–63] More broadly, occupational exposure to air pollution and noise affects quality of life, mental health, and job performance for millions of workers.[56,58,64,65] The economic cost of these work-related hazards is estimated to be multiple billions of dollars in high-income countries like the United States (US) and Australia.[66–68] Global statistics offer a big-picture view of the problem, but harm reduction requires accurate, comprehensive, and actionable data on individual exposures.

Infrequent workplace inspections, combined with the limited collection of personal exposure samples during inspections, limit the data available for comprehensive exposure assessments. Globally, resources for occupational exposure assessment, control, and management are very limited, impacting data availability and, ultimately, effective worker protection. Although institutions such as WHO, ILO, and national entities that supervise occupational health [e.g., the Health and Safety Executive (HSE) in the UK, the National Institute for Occupational Safety and Health (NIOSH) and the Occupational Safety and Health Administration (OSHA) in the US] publish country- and sector-level occupational exposure data, the underlying worker-level data are scarce, inconsistent, and vary in quality, leading to high uncertainty in population-level estimates. From 1984 to 2009, OSHA analyzed an average of ~44,000 personal samples per year including all hazard types, and the median number of samples collected per workplace inspection was between two and three.[69] The manufacturing sector in the US alone employs >8 million people, and there are >400,000 workers in just one high-risk set of occupations within this sector that includes welders, cutters, solderers, and brazers.[70] The total number of workers in high-risk sectors (e.g., construction, oil and gas, mining) is much larger. Overall, OSHA-driven efforts collect personal exposure samples for less than 1% of at-risk US workers, and even less for each hazard. Occupational exposure databases from other G7 countries [e.g., Messdaten zur Exposition gegenüber Gefahrstoffen am Arbeitsplatz (MEGA) in Germany, Canadian Workplace Exposure Database (CWED) in Canada, National Exposure Database (NEDB) in the UK, Système de Collecte des Données d'exposition Chimiques des Laboratoires des Caisses Régionales d'assurance Maladie (COLCHIC) and Système de COllecte des Laboratoires Accrédités (SCOLA) in France] report similar numbers of personal samples per year,[71–74] whereas such information is not readily available for low- and middle-income countries.

Employers, researchers, and other non-governmental actors also carry out exposure assessment work, but the extent of those efforts is hard to quantify given that the resulting data are generally not available from public databases and remain scattered/uncollated.

The typical workplace inspection sample size would have to expanded by ~10× (i.e., from 3 to 30) to capture the variability in exposures to air pollution and noise, which are typically lognormally distributed,[3,7,8,75] but there are major barriers to expanded personal exposure monitoring. The main limiting factor is the cost of equipment and specialized labor. In most cases, one sampler can only monitor one worker's exposure to a single hazard and each industrial hygienist (IH) can deploy approximately 8 instruments per day, so a large inspection requires several IHs.[76] Furthermore, if more than one type of exposure (e.g., aerosols, gases, noise) must be measured, an additional instrument per worker for each type of exposure is needed, and additional personnel are required to deploy the devices. For 30 workers (i.e., a sufficient sample size to estimate the 95th percentile of a lognormal distribution), measuring three types of exposure could require 90 instruments and ~10 IHs, costing upwards of $50,000 including the use of equipment, sample analysis, and consultant time. The burden on workers (i.e., discomfort associated with wearing monitoring equipment), loss of productivity, and the complexity of protocols required to obtain high quality data are additional challenges associated with personal exposure assessment.[76,77] Monitoring tools and procedures must be modernized to enable higher sampling throughput.

This work describes a full-scale deployment of the AirPen, a small wearable sampler for particulate matter (PM), volatile organic compounds (VOCs), and noise.[19] The AirPen was designed to reduce the total cost of personal monitor deployment and to make large sample-size personal exposure assessment campaigns attainable by integrating the functionality of multiple

samplers into a single, easy-to-deploy device. Moreover, the AirPen includes a suite of on-board diagnostic tools that streamline the sample quality assurance process. The main objective of this study was to measure the exposures of nearly all workers in a single large facility (~100 workers) in a single day. We carried out a sampling campaign at a furniture manufacturing facility in the US. We collected time-integrated samples, as well as a wide array of time-resolved sensor data. This work focuses on exposures to total aerosol mass, formaldehyde, and noise. All three of these hazards are relevant to the workplace in this study, are frequently encountered in the manufacturing industry, have established methods for exposure assessment, and have OSHA-defined occupational exposure limits (OELs).

Materials and methods

Study site and hazard selection

We partnered with a major furniture company to formulate the sampling protocols for a facility-wide campaign at one of their manufacturing and assembly factories in Georgia, USA. The size of the workforce at this facility varies based on the required production volume but typically ranges from 80 – 130 workers, which was adequate for our desired sample size. There are several different production lines within the ~15,000 m^2 facility (see descriptions of production lines in Appendix B section 2.1). The company's industrial hygienists identified relevant work hazards based on prior measurements, expert judgement, and exposure control programs already in place at this facility. To inform the protocol design, we collected preliminary data in different areas of the facility using stationary monitors (results in Appendix B section 1). To demonstrate the AirPen's multi-hazard capabilities, we selected one particle-phase and one gas-phase analyte for measurement. Total particulates not otherwise regulated (i.e., total aerosol mass), as defined by NMAM 0500 method,[78] was chosen as the particle-phase analyte based on the prevalence of

large particles (e.g., wood dust) and the lack of evidence for the presence of particularly hazardous dusts (e.g., crystalline silica, cobalt, vanadium). Formaldehyde was selected as the gas-phase analyte because the company had recently phased out multiple products that contained formaldehyde and wanted to assess workers' exposure to formaldehyde relative to established OELs. They anticipated that formaldehyde exposures would be low following implementation of the new control program, but measurements were required for corroboration. The company also expressed interest in measuring noise exposure given that third-party inspections had measured noise exposures close to OELs, necessitating the establishment of a hearing protection program.

Personal exposure monitoring

Personal exposures to total particulates, formaldehyde, and noise were measured using the AirPen. The design, laboratory validation, and field validation of the AirPen has been described in detail by Tryner et al.[19] The AirPen combines multiple environmental sensors with pumping technology to support time-integrated sampling of particles and gases simultaneously (Figure 5).

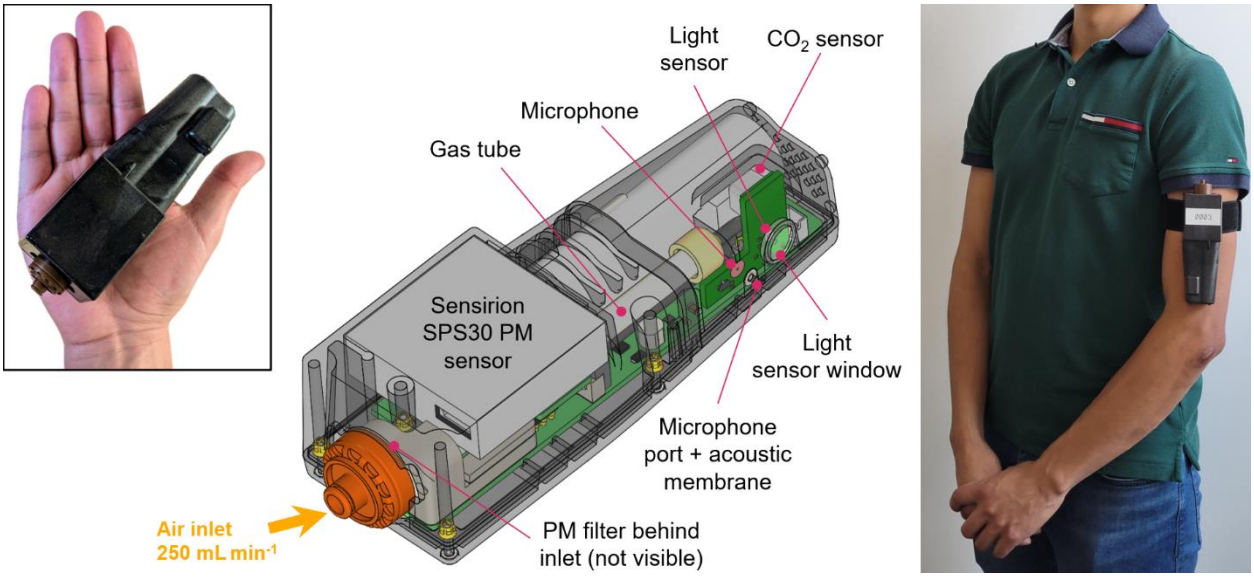

Figure 5: AirPen diagram and pictures. Center: a view of the fully assembled AirPen, with the housing shown as transparent so that the main components are visible. Top left: a photograph of the AirPen in the palm of a person's hand for scale. Right: a photograph of a person wearing an AirPen.

This device uses mass flow and temperature/humidity/pressure sensors to control the flow rate (250 mL min^{-1} default) through a 15-mm filter for PM sampling. A vapor adsorption cartridge is installed downstream of the filter to sample gas-phase pollutants. The AirPen was specifically designed to accommodate 6.35-mm-outer-diameter × 89-mm-long gas sampling tubes (e.g., stainless steel thermal desorption tubes or glass LpDNPH tubes). The flow rate through the tube is passively controlled by an orifice and continuously measured by a secondary mass flow sensor. Depending on the type of tube, 5% – 10% of the air that goes through the filter is directed to the gas cartridge (i.e., 12 – 25 mL min^{-1} under default configuration), and the rest bypasses the cartridge. Sensors within and outside of the PM and gas sampling manifold measure PM, total volatile organic compounds (tVOC), temperature/humidity/pressure, and sampling flow rates. Additionally, the AirPen records accelerometer, ambient light, and GPS data. Sensor data are logged on SD cards in 30-second intervals.

For this work, we designed a PM inlet that emulates, at the lower flow rate of the AirPen, the sampling efficiency of the simple and disposable sampler for inhalable aerosol described by L'Orange et al.[79] The device described by L'Orange et al. was designed to sample particles in accordance with the inhalable particulate mass (IPM) criterion, which requires a higher sampling efficiency for particles >30 μm than the 37-mm closed-face cassette (CFC) typically used for the total aerosol mass method (NMAM 0500).[80,81] Therefore, our method is meant to be more protective (i.e., our inlet design will slightly overestimate exposure relative to the CFC). For this study, we also modified the AirPen described by Tryner et al.[19] to incorporate a photoacoustic NDIR sensor (SCD41, Sensirion, Stäfa, Switzerland) for CO_2 monitoring and a MEMS analog microphone (IMP23ABSUTR, STMicroelectronics, Geneva, Switzerland) for noise monitoring. The microphones were calibrated with a 1-kHz tone in the 50 to 110 dBA range, using a Spark

703+ dosimeter (Larson Davis, Depew, NY, USA) as the reference (Appendix B section 9.2). Then, two calibrated AirPens were colocated with two Spark 703+ dosimeters and worn in various industrial settings to validate noise monitoring performance (Appendix B section 9.3).

Laboratory pre-deployment

AirPen settings and sampling parameters (e.g., metadata, flow rate, sensor modes) were configured prior to deployment using a mobile phone application so that each sampler could be shipped to the field in a "ready-to-use" state. A primary flow rate of 250 mL min^{-1} for the PM sample was selected. AirPens were pre-configured in "next start-up" mode. In this mode, the device saves settings and metadata in non-volatile memory and immediately starts a sample at next power-on. Prior to shipping, we installed 15-mm diameter PTFE membrane filters (PT15P-PF03, Measurement Technology Laboratories, Minneapolis, MN, USA) for PM sampling and ORBO™ LpDNPH glass tubes (20081-U, Sigma-Aldrich, St. Louis, MO, USA) for aldehyde sampling. We shipped 83 functional AirPens one week before the sampling campaign to ensure timely delivery. Nine additional units were loaded with identical sampling media, packed in the same manner, and shipped in the same cases to be used as field blanks. Additional details on sampler preparations are available in Appendix B (sections 3.1 and 3.2).

Field pre-deployment

Although the instruments were pre-configured, the field team arrived 2 days before the sampling campaign to have enough time to organize logistical details, check the condition of AirPens, and charge batteries. This last task was inevitable because rechargeable lithium-ion batteries must be shipped at 30% state of charge or less in the US.[82] The day before the deployment, a final inspection of the physical condition of the devices and sampling media uncovered a significant

number of broken LpDNPH glass tubes (Figure B8). To minimize the consequences of this mishap, the field team disassembled all AirPens with broken tubes, cleaned them thoroughly, and capped the tube inlets and outlets on the samplers. This work allowed those instruments to still be used for PM and noise sampling.

Participant recruitment

The exposure assessment campaign was one component of a larger project at this facility and participants were recruited at various times before the field campaign (details in Appendix B section 3.3). All participants provided informed consent and completed an intake survey. The only survey data used for the work presented here was the self-reported job category (i.e., department and/or production line). Most participants completed the intake requirements multiple days before the sampling day, but newly-recruited workers had to complete the consent process and the intake survey right before sampling, adding time (~15 minutes) and complexity to the deployment process (Figure B10). All efforts were made to include everyone who wanted to participate, if they met the inclusion criteria (i.e., at least 18 years old and employed by the furniture company). This study was approved by the Institutional Review Board at Colorado State University (protocol #2364).

Sampling

On the morning of the deployment, participants were fitted with an Airpen using elastic straps around their preferred arm (Figure B11), and the device was turned on to start the pre-programmed sample automatically. This process took about 1 minute per participant. Participants were asked to bring the AirPens back at the end of their shifts. The field team turned off the devices and provided a post-deployment survey. The AirPens were stored in the shipping cases at

the end of the day (protected with foam dividers, but not in individual sealed bags). The next morning, filters and gas tubes were removed from the devices and transferred to individual sealed containers. We removed the sampling media from the AirPens at this point to avoid sample contamination and breakage during shipping and to expedite analysis. Additional details on sampling and media handling are available in Appendix B (section 3.4).

Physical sample analysis

Collected PM mass was determined to the nearest 1 μg using an automated robotic weighing system (for both pre- and post-weighing).[83] The system uses a robotic arm (UR3, Universal Robots, Odense, Denmark) to weigh sample filters on an analytic microbalance (Mettler Toledo XS3DU, Columbus, OH, USA) within a humidity-controlled enclosure. A polonium-210 radiation source (2U500, NRD, Grand Island, NY, USA) is used to remove static charge before each weighing. Further details on the PM analytical method are available in section 4.1 of Appendix B.

LpDNPH tubes were analyzed using ultraviolet high-performance liquid chromatography (HPLC-UV) (1260 Infinity HPLC, Agilent, Santa Clara, CA, USA) using a method derived from EPA TO-11A. This method was developed and used previously to measure carbonyls emitted from biomass-fueled cooking stoves.[84–86] An extended description of the analytical procedures is available in Appendix B (section 4.2).

Noise level calculations

The analog signal from the microphone was measured by the internal analog-to-digital converter (ADC) on the AirPen's microcontroller (STM32WB5MMG, STMicroelectronics, Geneva, Switzerland) at 50,000 Hz and 14-bit resolution. ADC values were logged to the SD card and

transferred to a personal computer to calculate loudness levels. All data analyses from hereon were executed using R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). Fast Fourier Transform (FFT) was used to convert the time domain ADC values to frequency domain data, and A-weighting was subsequently applied to calculate loudness in dBA.[87] The individual calibration curve developed for each AirPen (Appendix B section 9.2) was used to adjust the raw calculated, A-weighted loudness values. Noise data processing and calculations are described in detail in section 9.1 of Appendix B.

Data and sample integrity

Instrument malfunction, operation errors, participants not wearing the sampler correctly, and other factors can lead to data/sample losses or introduce biases. Rigorous exposure assessment requires significant time for data quality controls.[88,89] We used AirPen sensor data and on-board diagnostics to automate most of the sample validation work. First, we removed all samples with runtimes shorter than four hours and samples with corrupted/missing AirPen log files. For noise, log files were analyzed to identify signs of missing or faulty microphones. Examples of these signs included steady ADC voltages close to ground (0 V), to supply voltage (3.3 V), or to reference voltage (1/2 of supply voltage). Flow rate and pump voltage data recorded by sensors on AirPens were used to identify flow control anomalies (details in Appendix B section 5). Samples with 5 or more minutes of anomalous flow control data were flagged as potentially anomalous samples. Time-resolved flow rate and pump voltage data from potentially anomalous samples were graphed, visually assessed, and compared to data from known faulty AirPens (Figure B13). If the visual assessment evidenced a trend consistent with device malfunction, the sample was corroborated to be anomalous and removed.

If a worker does not wear a sampler for the full sample duration, the resulting measurement might not be fully representative of that worker's occupational exposure.[90] AirPen accelerometer data were used to identify times when samplers were abnormally steady (details in Appendix B section 6), which would be indicative of the sampler no longer being worn. Samples with more than 20% of anomalous accelerometer data (i.e., 96 minutes of total anomalous accelerometer data for an 8-hour sample; not necessarily consecutive) were flagged as potentially anomalous samples. Time-resolved acceleration and step count data from potentially anomalous samples were graphed and visually assessed to confirm that the periods flagged by the algorithm were consistent with the sampler not being worn.

Left-censored data treatment

Several PM and formaldehyde samples had concentrations below the calculated limits of detection. Numerical values were available for those samples, so we applied different treatments to those data—truncation (i.e., removing samples below LOD), imputation, and keeping the values as reported by the laboratory—to see how distribution estimates and summary statistics were affected by each treatment. For imputation, we used Pleil's multiple-ordered value imputation method due to its effectiveness and simplicity, and because numerical values for samples below the LOD were available. Pleil showed that multiple-ordered value imputation produces better distribution estimates and summary statistics than data truncation, single imputation (e.g., replace all samples with LOD/2 or LOD/$\sqrt{2}$), and multiple random imputation, without the complexity of other methods such as β-substitution or Bayesian methods.[91] In multiple-ordered value imputation, the values below the LOD are replaced with specific values calculated to obtain a straight QQ-plot. The order can be determined using the laboratory results of the analyte of interest when available (i.e., our case) or a different correlated analyte.

Fitting distributions to empirical exposure data

Distributions were fitted using closed-form maximum likelihood estimation (MLE) via the *MASS* R package.[92] Histograms were used to select distribution type candidates. Goodness of fit was assessed by comparing the empirical and theoretical probability density and cumulative density graphs, as well as the Q-Q and P-P plots (i.e., visually).[93] Likelihood-based criteria [Akaike information criterion (AIC) and Bayesian information criterion (BIC)] were used to quantitatively compare fitted distributions. Applicable distribution parameters were calculated for each fitted distribution: mean (μ) and standard deviation (σ) for normal distributions, mean [$\mu(\ln x)$] and standard deviation [$\sigma(\ln x)$] of the log-transformed random variable for lognormal distributions, and scale (λ) and shape (k) for Weibull distributions. For lognormal distributions, the geometric mean and geometric standard deviation were calculated as $exp[\mu(\ln x)]$ and $exp[\sigma(\ln x)]$, respectively. The 95th percentiles were also calculated for all fitted distributions using the cumulative density functions. Upper confidence limits (UCLs) for the 95th percentiles were estimated using k factors for one-sided tolerance intervals (Eq. 1 and Eq. 2).[94,95]

$$UCL_{95\%} = GM + (GSD)^{k_{95\%,\varphi,N}} \quad for\ lognormal\ distributions \tag{1}$$

$$UCL_{95\%} = \mu + k_{95\%,\varphi,N} \cdot \sigma \quad for\ normal\ distributions \tag{2}$$

The $UCL_{95\%}$ is defined as a threshold such that the real 95th percentile is smaller than the $UCL_{95\%}$ at a ϕ confidence level. In equation 1, GM is the geometric mean, and GSD is the geometric standard deviation. In equation 2, μ is the arithmetic mean and σ is the standard deviation. In equations 1 and 2, $k_{95\%,\varphi,N}$ is the k factor for a 95th percentile at a ϕ confidence level and a sample size N.

Summary statistics

Basic descriptive statistics and metrics that could not be calculated from fitted distributions were obtained from the empirical data. This is the case for minimums, maximums, medians, interquartile ranges (IQR), and sample sizes of all variables, as well as geometric means and geometric standard deviations of normally- and Weibull-distributed variables.

Similar exposure groups (SEG) analysis

Production lines described in Appendix B section 2.1 were analyzed as SEGs. Groups were analyzed at two levels of granularity. First, Welch's two-sample t-tests were used to determine whether mean exposures were significantly different between non-factory (i.e., administration or shipping) and factory (i.e., the rest) workers. Then, one-way analysis of variance (ANOVA) and Tukey multiple pairwise comparisons were used to determine if mean exposures between production lines were significantly different. The differences in variances between production lines were assessed visually with boxplots.

Results

Sampling success rates

Sample validation is shown in detail in Figure 6. Most of the sample attrition was caused by the LpDNPH glass tubes that broke during shipping. This single incident reduced the formaldehyde sample size significantly (attrition = 47; 57%); it also forced us to disassemble, modify, and reassemble the AirPens last-minute. When the AirPens were reassembled, some microphone boards were connected incorrectly, reducing the noise sample size (attrition = 10; 12%). In addition, four samplers were not turned on correctly in the morning, and one sampler was not turned off at the end of the work shift.

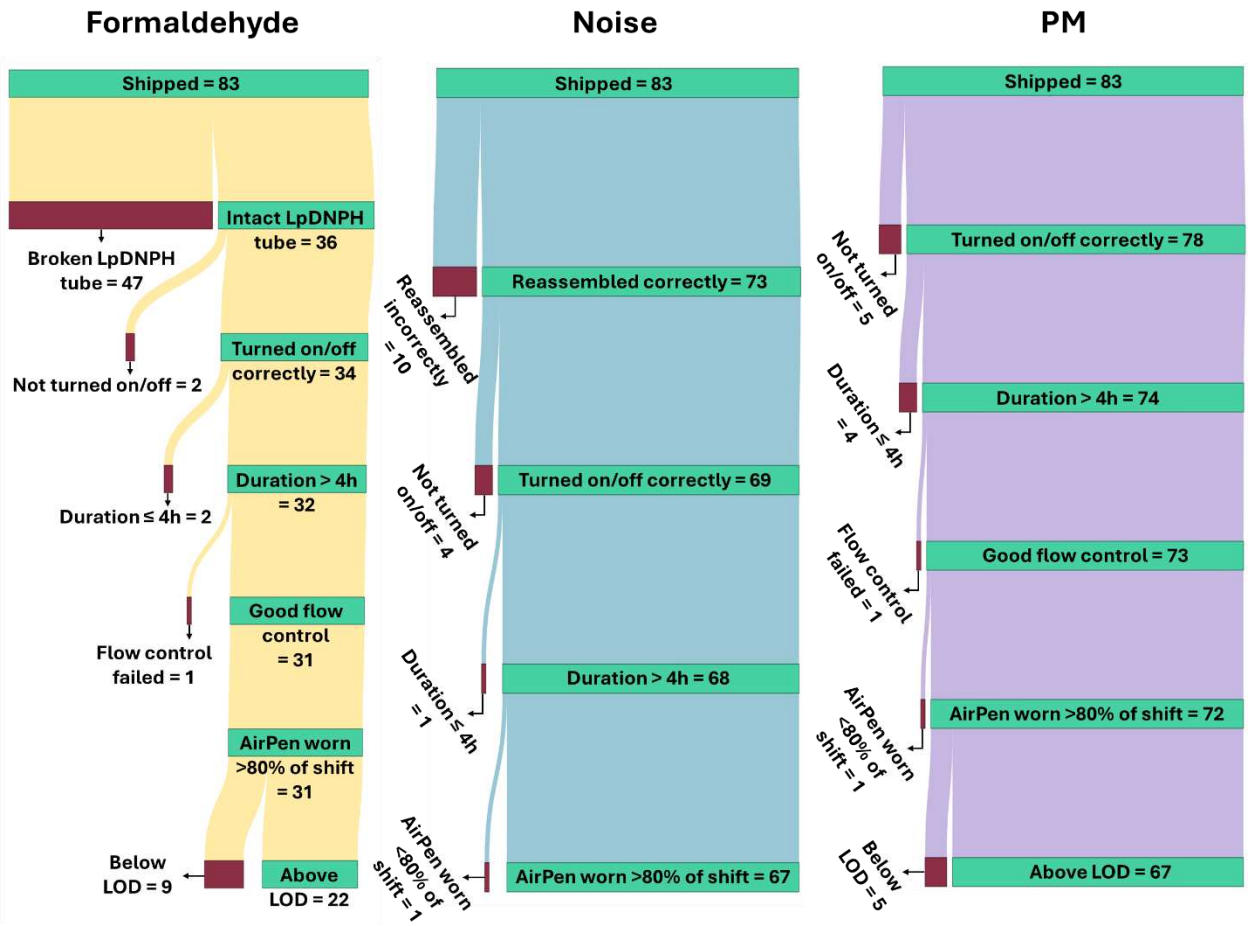

Figure 6: Sankey diagram of sample size attrition. Green nodes correspond to samples that met QA/QC criteria. Red nodes correspond to invalid samples. "Reassembled correctly/incorrectly" refers to the reassembly after the broken LpDNPH tubes were removed and samplers were cleaned.

Four additional AirPens produced short samples (< 4 hours) due to firmware issues (i.e., errors in the source code that were identified and fixed after this study). Five PM filters had masses below the LOD (equivalent to 86.7 µg m^{-3} for an 8-hour sample). Nine formaldehyde samples had peak areas below the LOD (equivalent to 3.1 ng or ~8 ppb for the median sampled volume). Detailed detection limit calculations are available in the supplement (Appendix B sections 4.1 and 4.2). Anomaly detection algorithms also detected a flow control issue (i.e., leak) on one sampler (see Figure B14). Three additional units were flagged as potentially anomalous by the flow analysis algorithms, but no problems were found when flow rate and voltage data were visually inspected

(Figures B15, B16, B17). Finally, analysis of accelerometer data detected one sampler that was only worn by the worker for ~5 hours out of the 8-hour shift (Figure B18). After applying all the validation criteria, 67 PM, 67 noise, and 22 formaldehyde samples remained, corresponding to success rates of 81% for PM, 81% for noise, and 27% for formaldehyde. If samples below the limit of detection are counted as valid, the success rates are slightly higher: 87%, 81%, and 37% for PM, noise, and formaldehyde, respectively.

Exposure to total particulates (PM)

Table 2: PM statistics. Distribution parameters, central tendency, dispersion, and upper percentile estimates. Statistics are disaggregated by left-censored data treatment.

Total Particulate Matter [µg/m^3]

	Below LOD treatment		
Parameter	*Truncation*	*Imputation*	*Data as reported*
Estimated from fitted distribution			
$\mu(\ln x)$ (std. err.) [ln(µg/m^3)]	5.501 (0.069)	5.437 (0.071)	5.430 (0.072)
$\sigma(\ln x)$ (std. err.) [ln(µg/m^3)]	0.5651 (0.0488)	0.5993 (0.0499)	0.6111 (0.0509)
GM	244.8	229.8	228.3
GSD [unitless]	1.767	1.828	1.850
95th pct.	620.22	615.7	623.7
$UCL_{95\%}$ (90% conf.)	720.1	716.9	728.4
$UCL_{95\%}$ (95% conf.)	755.5	752.6	765.4
Estimated from raw data			
Sample size	67	72	72
Minimum	84.3	52.0	42.9
Maximum	941.8	941.8	941.8
Arithmetic mean	289.0	276.1	275.6
Median	233.9	228.7	228.7
IQR	152.8	162.0	162.0

Particulate matter exposures appeared lognormal, in line with the existing literature (Figure 7a). Left-censored data treatment had a minor effect on the results (Table 2). The truncated dataset (i.e., samples below LOD removed) was slightly biased towards higher concentrations relative to the imputed and "as reported" datasets, but the distribution parameters (μ and σ) were within the standard errors from each other. The largest differences between datasets were observed around the left tails. The histograms showed clear differences below 200 μg m^{-3} (see Figure 7a), and the minimum concentrations differed by a factor of 2 between the truncated and "as reported" datasets. In this study, every sample was well below the 15,000 μg m^{-3} OSHA 8-hour permissible exposure limit (PEL) for total particulates not otherwise regulated.[96] Measured exposures and $UCL_{95\%}$ estimates were also below NIOSH recommended exposure limit (REL) of 1,000 μg m^{-3} for wood dust.[97]

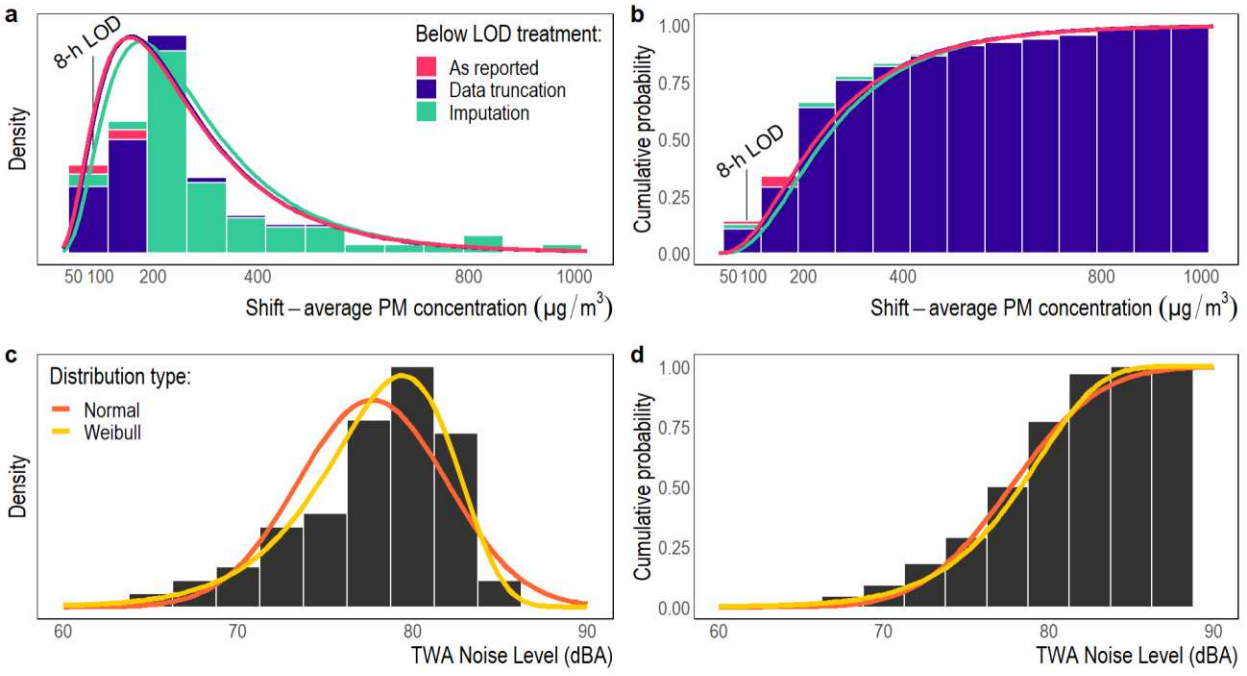

Figure 7: PM (top) and noise (bottom) exposure distributions. **a)** PM exposure distribution is shown as histograms and fitted lognormal probability density functions. Colors represent the left-censored data treatment. **b)** PM exposure distribution is shown as cumulative histograms and fitted lognormal cumulative distribution functions. Colors represent the left-censored data treatment as specified by the legend on panel a. **c)** Noise exposure distribution is shown as a histogram and two different types of fitted probability density functions. Colors represent

the fitted distribution type. **d)** Noise exposure distribution is shown as a cumulative histogram and fitted cumulative distribution functions. Colors represent the fitted distribution type as specified by the legend on panel c.

Exposure to noise

Less than 0.2% of noise measurements were below the microphone's LOD (i.e., < 50 dBA). Thus, we did not apply any left-censored data treatment (i.e., there was a single set of empirical measurements). Upon visual inspection, the histogram did not show a clear distribution type candidate. Both normal and Weibull distributions approximated the empirical data well (Figure 7c). The Weibull distribution had an Akaike information criterion (AIC) of 375 whereas the normal distribution had an AIC of 382 (lower means higher goodness of fit). The Bayesian information criteria (BIC) confirmed that the Weibull distribution fit the empirical data better (379 vs. 387); however, not all of the summary statistics used for occupational exposure assessment and epidemiology are applicable to Weibull distributions, and others are hard to estimate. Therefore, some parameters were only calculated using the fitted normal distribution and empirical data (Table 3). The arithmetic and geometric means were nearly identical at 77.7 ± 0.1 dBA. The standard deviation, GSD, and range (20.8 dBA) evidenced a compact distribution. Furthermore, the interquartile range was 5.55 dBA, so most of the measurements were concentrated in a narrow band. The small variability produced very compact confidence intervals. $UCL_{95\%}$ estimates were only 1.0 – 1.5 dBA higher than the normal distribution 95^{th} percentile estimate, and ~3 dBA higher than the empirical 95^{th} percentile. $UCL_{95\%}$ estimates were also within 0.25 dBA of the maximum measured noise exposure. The noise exposure results from this campaign were below the OSHA 8-hour PEL of 90 dBA[98] by at least 3.9 dBA. Nonetheless, one measurement and the $UCL_{95\%}$ estimates were above the OSHA action level of 85 dBA, which means that the existing on-site hearing conservation program is required (but engineering controls are not).[98]

Table 3: Noise statistics. Distribution parameters, central tendency, dispersion, and upper percentile estimates. Statistics are disaggregated by data source used to estimate the parameter.

Noise TWA [dBA]

	Estimate source		
Parameter	*Normal fit*	*Weibull fit*	*Raw data*
μ (std. err.)	77.77 (0.52)	N/A	77.77
σ (std. err.)	4.258 (0.368)	N/A	4.290
Scale λ (std. err.)	N/A	79.65 (0.45)	N/A
Shape *k* (std. err.) [unitless]	N/A	22.78 (2.18)	N/A
GM	N/A	N/A	77.65
GSD [unitless]	N/A	N/A	1.058
95th pct.	84.77	83.58	82.87
$UCL_{95\%}$ (90% conf.)	85.90	N/A	N/A
$UCL_{95\%}$ (95% conf.)	86.26	N/A	N/A
Minimum	N/A	N/A	65.24
Maximum	N/A	N/A	86.06

Exposure to formaldehyde

Given the low sampling success rate for formaldehyde, only a brief description is presented here (extended results are included in Appendix B section 7). Based on the histograms, the distribution of formaldehyde exposure looked like a mixture of two lognormal distributions (i.e., bimodal lognormal), but the valley between the two modes could also be caused by missing data (Figure B19). Given that a simple normal or lognormal distribution could not be fitted, we calculated the summary statistics using the empirical data (Table B1). The arithmetic mean ranged from 13.8 to 17.8 ppb, and the geometric mean ranged from 13.5 to 17.1 ppb, depending on the left-censored data treatment. Standard deviation ranged from 4.3 to 7.4 ppb, and GSD ranged from 1.3 to 1.9. Notably, even the highest measured formaldehyde concentration (25.6 ppb) was lower than OSHA's 8-hour PEL of 750 ppb by a factor of ~30. About a third of the formaldehyde samples were below the LOD.

Similar exposure groups (SEG) analysis

So far, we have presented results for the entire group of workers. In this section we analyzed the differences between and within departments or production lines to determine if those self-reported job categories could be used as SEGs. Welch's tests determined that the means of non-factory (i.e., administration and shipping) and factory workers (i.e., the rest) were different at a significance level of 0.05 for PM [difference in means (95% CI): 155.9 μg m^{-3} (77.8 – 240.4 μg m^{-3})], noise [difference in means (95% CI): 8.6 dBA (6.5 – 10.9 dBA)], and formaldehyde [difference in means (95% CI): 5.7 ppb (0.7 – 10.8 ppb)]. Boxplots also show that shipping and administration PM and noise exposures were consistently lower than factory worker exposures (Figure 8). Differences between factory production lines were less clear. For PM, the one-way ANOVA test did not reveal differences in means between factory production lines at any reasonable significance level (p-value = 0.515, DoF = 4). The same was true for noise (p-value = 0.135, DoF = 4) and formaldehyde (p-value = 0.595, DoF = 4). Tukey multiple pairwise comparisons did not reveal differences in means between any pair of factory production lines for any hazard (all p-values > 0.12). Results of all tests and assumption diagnostics are available in Appendix B (Section 8). Differences in variances can be assessed qualitatively using the boxplots and individual points in Figure 8. For PM, the machining line had a much narrower distribution compared to other factory lines. For noise, the final assembly line was the most compact, and its mean was the lowest of all factory lines. Variances of other factory lines were larger and closer to the population variance.

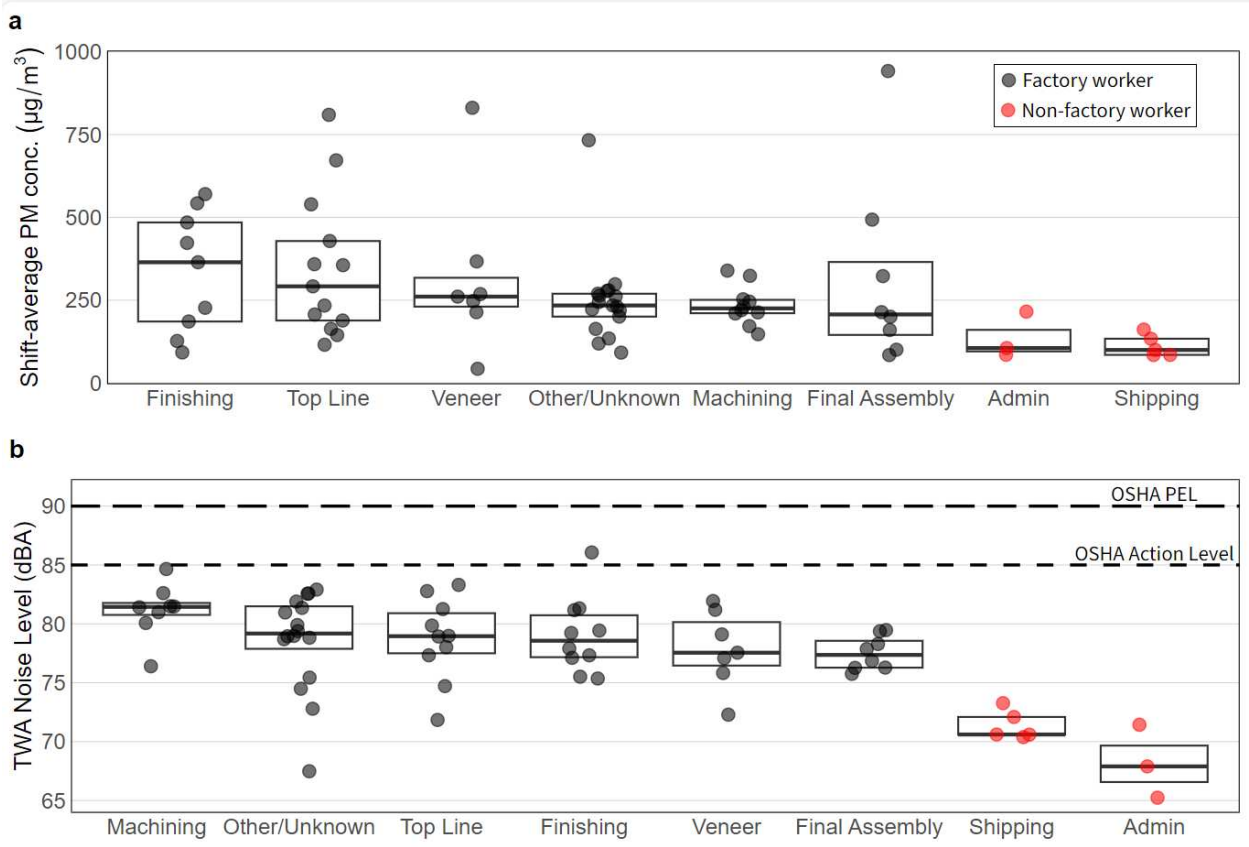

Figure 8: Exposure distributions by self-reported production line or department. Boxplots represent the 25th, 50th, and 75th percentiles for each SEG. Dots correspond to individual exposures. **a)** PM panel includes samples below the LOD. **b)** Noise panel shows the OSHA 8-hour action level (85 dBA) and PEL (90 dBA). Production lines and departments are ordered from highest to lowest median exposure (left to right).

Discussion

Enforcement of worker protection standards, as well as epidemiological research, often relies on estimates of exposure distribution statistics rather than complete primary exposure data. The latter is nearly unattainable in most cases, but sufficient sample sizes are crucial for accurate estimation of distribution statistics. Specifically, enough samples must be collected to estimate the 95th (or other upper) percentile to be compared against an occupational exposure limit (OEL).[95] For occupational epidemiologic studies, a reliable estimate of the central tendency of a given exposure is required (e.g., arithmetic or geometric mean).[99] In both cases, the sample size required to accurately estimate the distribution parameters depends on the distribution type and

the variability within the group. Occupational exposure to airborne pollutants often follows a log-normal distribution with high heterogeneity.[5–9] Noise exposure tends to follow a normal or log-normal distribution.[100–102] Hewett used computer simulations to calculate the sample sizes required to estimate the true arithmetic and geometric means of lognormal distributions within a certain accuracy. At least 20 samples are necessary for estimating the geometric mean (GM) of a log-normal distribution within ±20% of the true GM.[103] Selvin et al. noted that large sample sizes are required to identify "acceptable" environments when upper tolerance limits (UTL) are used to assess compliance of lognormally distributed exposures with 95% confidence and 95% coverage.[104]

Compliance with worker protection standards can also be evaluated using distribution-agnostic, non-parametric approaches. For example, the maximum risk subgroup strategy aims to sample at least one individual in the *high exposure subgroup* (e.g., top 10%), irrespective of the type of distribution.[105] The US National Institute of Occupational Safety and Health (NIOSH) sampling strategy manual includes equations to calculate the required sample size to obtain a high probability of sampling a *high-risk employee*. Equations are derived from the theory of sampling without replacement. For small groups ($N \leq 7$), every single worker must be measured to get a confidence of 90% of sampling at least one individual in the top 10%. At the same confidence level and exposure percentile, a group of 20 people requires 13 samples, and a group of 50 requires 18 samples. In all cases, the typical exposure assessment sample size (i.e., $2 \leq n \leq 5$) is insufficient by a factor of 5 – 10.

This study demonstrates a significant advancement in the feasibility of large-scale, multi-hazard exposure assessments and offers an adaptable solution to overcome historical limitations in occupational exposure data collection. The outcomes of this exposure assessment campaign

show how key barriers, such as logistical complexity and resource intensity, can be addressed through the integration of innovative technology and efficient deployment strategies. By simplifying the sampling process and reducing the burden on participants, the AirPen fosters greater participation and support from workers and employers alike, demonstrating its potential for broader application in occupational health research.

In this campaign, recruitment exceeded our expectations, and all the available samplers (n=83) were quickly assigned to workers who were interested in participating. The company was accommodating and interested in this work before, during, and after the deployment. Our team was allowed to recruit workers from all production lines freely. Workers were not intimidated by the sampling burden, and they were able to perform their usual duties. Loss of productivity was minimized by the quick-start functionality of the AirPen. Most important, the ability to collect facility-wide exposure data (for nearly every worker) dramatically reduced the uncertainty with predicting exposure distributions and estimating risk. Assessing the temporal variance and trends would add another layer of complexity that was out of the scope of this work. Factory-wide longitudinal studies could offer an unprecedented depth of knowledge about exposure.

PM exposure measurements demonstrated consistency with a lognormal distribution, and the treatment of left-censored data had negligible effects on the exposure assessment outcomes. For all treatments, the 95th percentile and its upper confidence limits, which are crucial for occupational compliance and worker protection, were within 2% of each other. The "as reported" dataset was marginally more protective. Notably, the $UCL_{95\%}$ estimates (717 – 765 μg m^{-3}) were lower than the maximum measured exposure (942 μg m^{-3}), indicating that we captured outliers that might otherwise be missed with a smaller sample size. To further illustrate the importance of sample size, using the same GM and GSD we measured, the $UCL_{95\%}$ estimates would be in the

1,200 – 1,300 µg m^{-3} range for a sample size of 10, and upwards of 2,500 µg m^{-3} for a sample size of 5 because the 95th percentile confidence interval width is a function of sample size.[94] Estimates generated from these smaller samples would change the exposure judgment relative to the NIOSH REL. This finding underscored the importance of robust sample sizes in occupational exposure studies, where smaller datasets could lead to overestimated confidence intervals and potentially flawed exposure judgments. Measures of central tendency, which are particularly relevant for epidemiological studies and long-term effect analysis, exhibited slight variations depending on the left-censored data treatment method. Differences were even larger at the lower end of the exposure range, showing the importance of LOD for relatively low concentrations. Although uncertainty at lower concentrations could be reduced by sampling at higher flow rates to collect more mass, these lower exposures remain a minor concern for worker protection and the most common applications of exposure assessment.

The noise exposure results from this study were particularly notable, as both the maximum and the $UCL_{95\%}$ estimates fell within the narrow band between the OSHA action level and PEL. If the uncertainty had been greater, as would be the case with a smaller sample size, these compliance indicators could have shifted below the action level or above the PEL, potentially resulting in insufficient worker protections (e.g., cancellation of a hearing protection program in the case of underestimation) or unnecessary investments in costly engineering controls (in the case of overestimation). The maximum measurement and the $UCL_{95\%}$ estimates were nearly identical (within 0.25 dBA), demonstrating reliability of the AirPen measurements. Therefore, using the maximum or the $UCL_{95\%}$ would lead to the same exposure judgement. Such agreement would be highly unlikely with a small sample size given that confidence intervals of the 95th

percentile would be much wider, reducing the likelihood of accurately capturing exposures in the top 5%.[105]

The limited number of formaldehyde samples collected in this study constrained the depth of analysis possible. This outcome underscored the need for optimized sampling protocols and improved reliability in future campaigns. Nonetheless, the low formaldehyde exposures measured across the factory provided strong evidence supporting the effectiveness of the program that replaced formaldehyde-emitting products with safer alternatives. When exposures are significantly below OELs, the increased uncertainty associated with smaller sample size becomes less critical for the purpose of assessing compliance with worker protection standards. However, for a more comprehensive understanding of workplace exposures, particularly for a carcinogen like formaldehyde, a larger sample size remains desirable. Collecting 20 – 30 formaldehyde samples in a single day was still a noteworthy achievement, particularly given formaldehyde's status as a known carcinogen and the challenges associated with its measurement. Future iterations of the AirPen should address reliability issues associated with sampling onto media in glass tubes.

Collecting exposure data for workers across all production lines allowed us to evaluate the applicability and utility of similar exposed groups (SEGs) for efficient exposure assessment and worker protection. Exposure assessment and worker protection can be more efficient when SEGs are considered. If workers within a group have very similar exposures (i.e., low intra-class variability), fewer samples might be required. If different groups have significantly disparate exposures (i.e., high inter-class variability), protection strategies can be tailored for each group. This approach helps prioritize monitoring efforts where they are most needed, ensuring that higher-risk groups receive adequate attention. However, accurately classifying workers into

appropriate SEGs remains a challenge that requires detailed analysis of exposure variability and patterns. In this study, the decision to treat factory and non-factory workers as separate SEGs was well supported, as their mean exposures were found to differ significantly. Non-factory workers had consistently lower exposures and do not require the same level of monitoring or protection. This distinction highlights how SEGs can guide targeted monitoring and protective measures to reduce unnecessary efforts for low-risk groups. Conversely, the analysis showed that the different production lines within the factory had similar mean exposures, making them unsuitable candidates for SEG-based differentiation apart from a couple of particular cases. For instance, the machining line had a low PM variance, suggesting it could require reduced PM exposure assessment efforts compared to other lines. Similarly, the final assembly line had low noise variance and a low mean exposure to noise, so it could be treated differently (e.g., the hearing protection program might not be required for people working exclusively in that line). The significant overlap in exposure distributions across the remaining factory lines does not justify distinct treatment or exposure assessment strategies for these groups.

The field team faced several challenges that impacted the success rates for each sample type (Figure 6). The incident with the broken LpDNPH glass tubes informed later projects; since the sampling campaign described herein, the AirPen design has been modified to accommodate plastic DNPH cartridges for aldehyde sampling. The rapid pace of deployment also introduced minor human errors in AirPen assembly and operation, highlighting the need for design improvements that reduce the likelihood of incorrect assembly, improved field protocols, or enhanced operator training to reduce such occurrences. Despite these obstacles, identifying weak points of the AirPen technology was valuable and informed later improvements in firmware and sampling logistics. Data logs of faulty AirPens were used to fix the firmware bugs that could not

be identified during smaller-scale studies. Other factors that reduced the number of valid samples were difficult to control or inherent to exposure measurement. For example, lower limits of detection could have been achieved by increasing sample flow rates, but this decision would have required *a priori* knowledge of the concentration range at the site. All in all, collecting more than 65 valid samples (≥ 80% success rate) for two different hazards (i.e., PM and noise) in a single day represented a significant improvement over traditional occupational exposure assessment campaigns, which have typically collected only two to five samples. Moreover, collecting 31 formaldehyde samples in parallel with the PM and noise samples was a remarkable achievement, even if the success rate was lower than expected. The success of our field sampling effort was strengthened by the comprehensive sample validation framework that we developed, which was only possible to implement at this scale using robust on-board diagnostic electronics and anomaly detection algorithms.

Conclusions

Comprehensive exposure assessments that include a sufficient number of individuals in a population are essential for making judgments supported by robust and reliable data. Currently, most vulnerable populations worldwide are not monitored—leaving public health studies, lawmaking, and enforcement efforts reliant on indirect or highly uncertain measures of exposure. This study demonstrates that technological innovations can address key barriers in exposure assessment, such as cost, time constraints, and participant burden. These findings underscore the importance of iterative development and rigorous validation to ensure the practical applicability of new methods in diverse occupational settings. This work provides an overview of the in-depth analyses possible when the majority of individuals in a group are included in exposure assessments. While the analyses presented here are not exhaustive due to space constraints,

further findings will be shared in future publications. Moreover, the campaign described here represents only a one-day snapshot of exposure, highlighting the need for longitudinal studies to capture temporal variability. Temporal variability is a crucial factor outside the scope of this project that needs to be studied, considering that exposure measurement campaigns are not only small in sample size, but also infrequent. Future development and case studies using the AirPen will focus on addressing the current limitations of the technology to further enhance its reliability and applicability.

Continued research and development on the AirPen technology can lead to samplers that are not only cheaper, more compact, and as reliable as established instruments, but also much more user-friendly for the public and less reliant on expert setup and supervision. Such multi-hazard monitors with on-board diagnostics could support exposome studies that have been too resource-intensive to be scalable in a variety of occupational and non-occupational settings. Our work demonstrates some of the steps needed to achieve that breakthrough and marks a path forward.

CHAPTER 4. A LOW-COST, COMMUNITY-BASED PLATFORM FOR MONITORING AMBIENT CONCENTRATIONS OF HAZARDOUS AIR POLLUTANTS

Chapter overview

Exposure to hazardous air pollutants (HAP) is linked to elevated cancer risk, immune system damage, respiratory problems, reproductive issues, and other severe health effects. Regulatory HAP monitoring is expensive, often performed with limited spatial resolution, and impeded by power outages or extreme weather events. Unfortunately, such extreme events often coincide with times when HAP monitoring is needed most. This work describes the development of a low-cost, battery-powered, community-based sampling platform for HAPs. The sampler was designed to provide reliable and robust measurement data by monitoring key quality assurance/control parameters during sampling, namely: time, location, battery power, sample flow rate, temperature, humidity, and pressure. We partnered with and trained local volunteers to deploy the technology for ambient formaldehyde sampling (following EPA method TO-11A) in Houston, TX. We obtained 28 fully validated samples over a six-week period, ranging in concentrations from 1.27 to 7.30 ppb (TWA = 2.36 ppb); these values were consistent with concentrations reported by nearby regulatory monitors. Non-cancer effects are unlikely at the levels reported in this work, but the contribution to cancer risk is non-negligible. With minor technological improvements, and refined user training and communication protocols, our monitoring platform shows promise for producing reliable, community-led data with higher spatial coverage and at lower cost than established monitoring alternatives.

Introduction

Exposure to hazardous air pollutants (HAP) is linked to elevated rates of cancer and severe damage to the immune system, the respiratory apparatus, the reproductive organs, and other systems. Despite their toxicity, HAPs are less monitored and regulated than criteria air pollutants (e.g., particulate matter, ozone, and nitrogen dioxide).[106] This imbalance in oversight leaves critical knowledge gaps that hinder our ability to protect public health. Federal regulations in the United States establish emission control requirements for major sources of HAPs.[107,108] EPA runs programs to measure outdoor concentrations of high-risk HAPs in some urban areas, and to inventory major sources, but nationwide allowable ambient limits (AAL) have not been established and federal monitoring requirements are modest.[109,110] Just one to three HAP monitoring stations are required in each state depending on their specific demographic and air quality conditions.[111] That leaves the control of ambient HAP concentrations in the hands of state and local governments, leading to inconsistent enforcement and monitoring. Consequently, monitoring for HAPs remains sparse and uneven, creating large blind-spots in our understanding of where and when air toxics pose the greatest risk.

Measuring volatile organic HAP concentrations is resource intensive. EPA has published 17 different methods for the determination of toxic organic pollutants in ambient air.[112] Although these methods vary in their required sampling media and operational parameters, they typically require a controlled (or known) volumetric flow of air into a canister, or through a cartridge containing adsorbent material or a reagent. Analysis of collected samples requires specialized instruments (e.g., gas/liquid chromatography, mass spectrometry), and calibration standards. A small number of quasi-real-time monitoring methods is also available.[112] One key challenge with HAP sampling is maintaining adequate standard operating procedures that ensure data quality

and traceability. Therefore, samplers often integrate features such as automated leak detection, logging of flow parameters, and out-of-tolerance condition monitoring. To protect sample integrity, personnel must also handle, transport, and store samples properly.[113,114] The cost, technical complexity, and training requirements of these established methods place them out of reach for most community-based monitoring efforts, leaving local stakeholders without the tools they need to generate reliable data.

Government-led monitoring has limited spatial resolution (i.e. zero to several stations per city) and can be affected by events like power outages and extreme weather.[115–117] Extreme weather events can lead to the release of unusually large amounts of HAPs or precursors due to infrastructure damage, resulting in monitors being offline when needed most. In Houston, TX, where our study took place, tornadoes, hurricanes, and flooding are common, especially in summer and fall. The Texas Commission on Environmental Quality, responsible for regulatory air monitoring in the state, shuts off its stations to protect the instrumentation and infrastructure during extreme weather events.[117] For example, in response to Hurricane Harvey (2017), local, state, and federal monitors were offline for days even as companies reported 319 tons of air toxics released, a tank failure emitted an estimated 12.5 tons of HAPs, and fires at a chemical plant combusted 160 tons of organic peroxides.[118] These dual challenges, namely, heightened emissions due to infrastructure failures and surveillance gaps when networks go offline, underscore the need for a streamlined, weather-resilient workflow that enables independent, community-based sampling of HAPs when needed.

In this work, we present the field demonstration of a low-cost, battery-powered, user-friendly platform for monitoring formaldehyde (a HAP with known carcinogenic effects)[107,119] and other carbonyl concentrations, carried out in Houston, Texas during the summer of 2024. We

developed a weather-resistant stationary version of the AirPen, a compact monitor for particulate matter (PM) and volatile organic compounds (VOC).[19] Houston was chosen for its major industrial HAP sources, elevated formaldehyde risk, and active community monitoring efforts: neighborhoods along the Ship Channel (e.g., Fifth Ward, declared a cancer cluster in 2019[120]) exhibit higher-than-normal cancer rates; the National Air Toxics Assessment identifies formaldehyde as the top cancer-risk contributor in nearly every Harris County tract; and the Houston Health Department recorded annual average formaldehyde concentrations as high as 2.27 ppb in 2019–2020 (≈1 in 50 000 extra cancer risk).[121,122] Our primary goal was to demonstrate that an independent, community-deployable platform could produce reliable carbonyl measurements under real-world conditions. To achieve this, we: (i) evaluated community-based deployment feasibility and sample integrity; (ii) benchmarked AirPen formaldehyde measurements against reference monitors in the study area; (iii) assessed spatial variability of carbonyl concentrations across multiple Ship Channel neighborhoods.

Materials and methods

Low-cost carbonyl samplers and sampling media

The AirPen was originally designed as a wearable personal exposure monitor for particulate matter and volatile organic compounds, as described in detail by Tryner et al.[19] The AirPen simultaneously collects low-cost sensor data and physical samples. An array of sensors (e.g., mass flow, temperature, pressure, humidity) is used for active flow control and quality control/assurance. For VOCs, the original design collected physical samples into thermal desorption tubes as required by EPA method TO-17. For this work, we modified the sampling train to accommodate Supelco BPE-DNPH cartridges (54278-U, Sigma-Aldrich, St. Louis, MO, USA). These BPE-DPNH cartridges are suitable for EPA method TO-11A for the determination

of formaldehyde concentrations in ambient air. The BPE bed acts as a built-in ozone scrubber that is not affected by high humidity[123]. A custom stainless-steel adapter and a commercially available Luer fitting were used to connect the BPE-DNPH in place of the thermal desorption tube (Figure 9a).

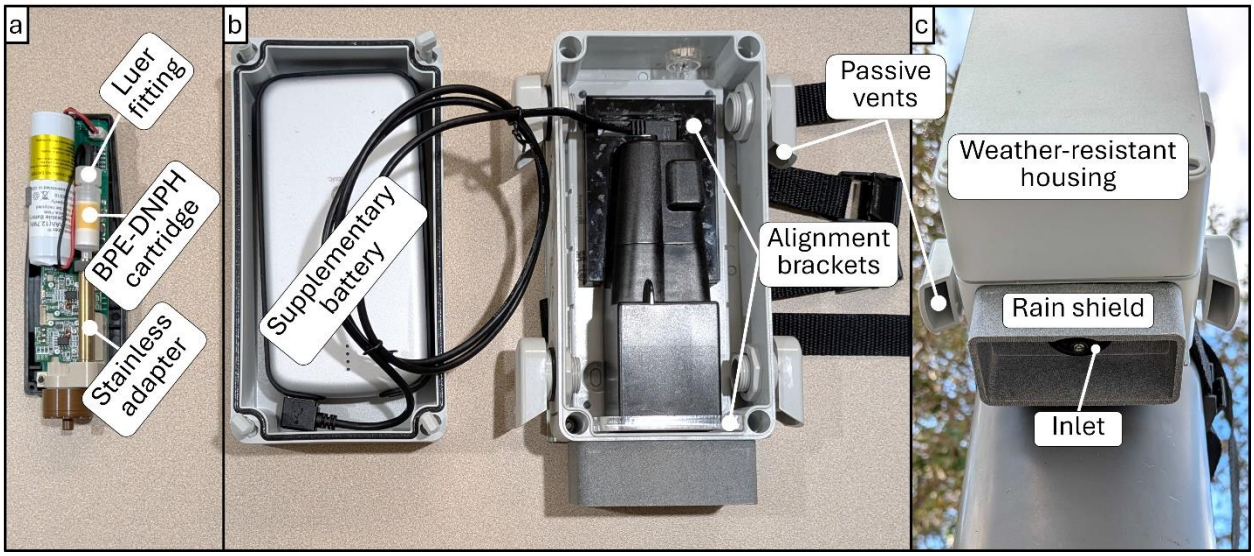

Figure 9: a) Photograph of the internal components of the AirPen, including the BPE-DNPH cartridge and the corresponding adapters. **b)** Photograph of the AirPen and the supplementary battery (connected with a cable) inside the weather-resistant housing. **c)** Photograph of the sampler mounted on a pole.

The AirPen runtime (12 to 16 hours) and flow rate through the gas tube (10 to 20 mL min^{-1}) were originally tailored for the high concentrations and short sampling durations of occupational exposure assessment. For the lower concentrations typically observed in ambient air (e.g., single-digit ppb for formaldehyde), a higher sample volume (e.g., > 50 L for low-ppb concentrations) was necessary to collect sufficient mass for analysis. A 72 W h battery pack (V75, Voltaic Systems, Brooklyn, NY, USA) was connected to the micro-USB charging port on the AirPen to supplement the 11 W h internal battery and extend the runtime (Figure 9b). The lower pressure drop of the BPE-DNPH cartridge, relative to thermal desorption tubes, allowed for higher flow

rates (i.e., 20 to 40 mL min^{-1}). The extended runtime time of 36 to 72 hours and increased flow rates made higher sample volumes of 50 to 150 L possible.

A custom weather-resistant housing was built to mount the AirPen outdoors securely and to protect it from the elements (Figures 9b and 9c). To replace the sampling media, users had to open the housing and detach the AirPen, which was held in place by alignment brackets and Velcro (Figure 9b). The AirPen uses a twist-cap mechanism mounted over the inlet as the on/off switch. Therefore, the inlet had to be accessible from the outside. With the AirPen mounted in the weather-resistant housing, the inlet faced down and stuck out enough for the user to operate the switch mechanism. A rain shield was installed around the inlet to prevent water ingress (Figure 9c).

Users controlled the devices through Bluetooth using an iOS application. All sensors (i.e., PM, total VOC, noise, CO_2) were turned on, except for the GPS, which was not used to save battery. The flow rate through the inlet is actively controlled by the sampler and was set to the default value of 250 mL min^{-1}. The flow rate through the BPE-DNPH cartridge is a small fraction (~8 – 16%) of the total flow rate through the inlet, passively controlled by an orifice in the manifold, and continuously measured by a mass flow sensor. Duty cycle was set to 100%. The sample duration was set to 72 hours.

Community-based field demonstration

Seven organizations (see section 1 in Appendix C) from different *super neighborhoods* (i.e., official territorial division in Houston) volunteered to collect samples in their communities (see map on Figure 10). The communities have HAP sources (e.g. refineries, other oil and gas facilities, plastic manufacturers, major highways etc.) nearby, and several of them are located

along the Houston Ship Channel (a hotspot for HAP sources). Eleven samplers were distributed across the different organizations (one or two per organization), and each sampler was assigned to a host who was responsible for collecting samples at their place of residence.

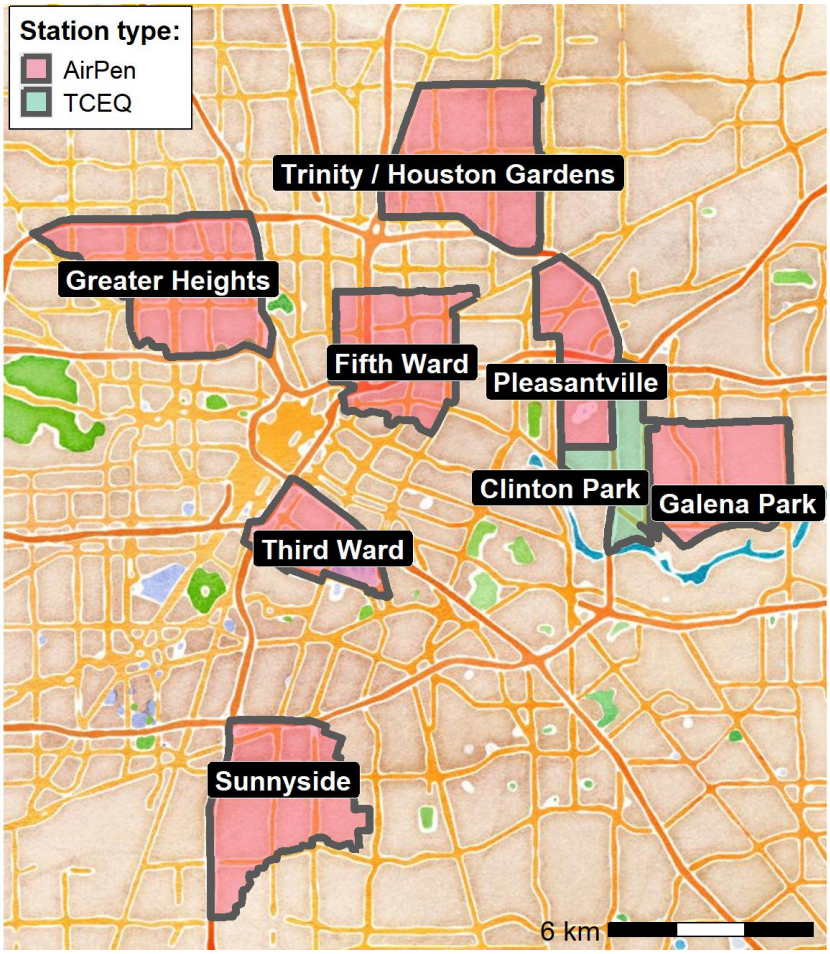

Figure 10: Map of the super neighborhoods where hosts installed samplers (in pink), and where the Texas Commission on Environmental Quality collects carbonyl samples at a regulatory station (in green). The Houston Ship Channel is shown in blue (south of Clinton Park and Galena Park). Orange lines represent highways and yellow lines represent major roads.

An in-person training session took place in July 2024, approximately three weeks before the start of the sampling campaign. During this session, hosts received kits that included the sampler, tools, a wall charger, a user manual, and BPE-DNPH cartridges for 6 weeks (plus one blank). Hosts who attended the training session had the opportunity to start a practice sample under the

guidance of the research team. Hosts who could not attend the session received the kit and were offered remote support as needed. Sampling media handling and blank collection procedures were covered in training and remainders were sent via mobile messaging at the end of each week. Hosts were asked to install cartridges and start a new sample weekly, preferably on Saturday or Sunday. After the sample ended by battery depletion (typically midweek), they were asked to remove the cartridge and store it in the original bag, sealed, and refrigerated. Batteries had to be recharged for at least 12 hours before the next sample could be started. Hosts were also asked to ship samples to Colorado State University for analysis weekly. The sampling campaign started on August 10, 2024 (i.e., during the peak of hurricane season in Houston) and lasted 6 weeks.

Physical sample analysis

BPE-DNPH cartridges were analyzed using ultraviolet high-performance liquid chromatography (HPLC-UV) (1260 Infinity HPLC, Agilent, Santa Clara, CA, USA). The analytical method, derived from EPA TO-11A, was originally developed to measure carbonyls emitted from biomass-fueled cooking stoves.[84–86] The HPLC-UV detector generates peak areas in absorbance units for a range of carbonyls, including formaldehyde, methacrolein, and propionaldehyde. We subtracted the mean reported peak area of the field blanks from the peak area reported by the instrument for each sample to account for bias introduced during handling and shipping. Blank corrected peak areas were transformed into equivalent carbonyl masses using calibration curves. The calibration curves were obtained using a mixed carbonyl standard solution in DNPH (CRM47285, Sigma-Aldrich, St. Louis, MO, USA). The carbonyl mass was divided by the volume of air sampled through the BPE-DNPH cartridge to obtain the average concentration in

air for each sample. Additional details on the physical sample analysis are available in section 2 of Appendix C.

Data and sample integrity

We used AirPen sensor data to validate samples. Samples with runtimes shorter than 12 hours and samples with corrupted or missing AirPen log files were removed. Anomaly detection algorithms based on flow rate through the inlet, flow rate through the BPE-DNPH cartridge, and pump voltage data were used to identify leaks, clogs, and other flow control issues (details in Appendix C section 3). Samples with more than 15 minutes of anomalous flow control data were flagged as potentially invalid. A visual assessment of the time-resolved flow rate and pump voltage data was conducted to confirm the existence of a fault and determine its nature. Samples were confirmed to be anomalous and removed from further analysis if the visual assessment of the time-resolved flow rate and pump voltage data showed a consistently anomalous trend as opposed to a small number of short anomalous events.

Regulatory air monitoring data

We downloaded HAP data from the Texas Air Monitoring Information System (TAMIS) to contextualize our measurements. TAMIS is managed by the Texas Commission on Environmental Quality (TCEQ) and is publicly accessible online.[124] The TCEQ station at Clinton Park monitors HAP concentrations and is adjacent to two of the neighborhoods where our samplers were installed (see Figure 10). TCEQ collects a 24-hour DNPH sample per week at Clinton Park and analyzes it using HPLC-UV to quantify carbonyl concentrations.[125] The proximity of the station to our sampling locations and the similarity in analytical methodologies gave us the opportunity to use the Clinton Park data as a reference for qualitative analysis.

Formaldehyde ambient concentration reference values and levels

No federal or state has established a binding ambient-air quality standard for formaldehyde; instead, agencies publish nonbinding screening or reference levels for health-based permitting and risk-assessment purposes. TCEQ uses a chronic noncarcinogenic effect screening level (ESL) of 2.7 ppb for air permit reviews, and a chronic noncarcinogenic reference value (ReV) of 8.9 ppb for air monitoring data evaluation. The carcinogenic ESL and ReV levels obtained by TCEQ are higher, so the more protective noncarcinogenic levels are used.[126] EPA's Integrated Risk Information System (IRIS) health assessment obtained a noncancer inhalation reference concentration (RfC) of 5.7 ppb, which is an estimate of a continuous exposure that is likely to be without risk of harmful noncancer effects over a lifetime. For cancer risk, EPA's IRIS toxicological review calculated an inhalation unit risk (IUR) of 1.1 x 10^{-5} per µg m^{-3} (1.35 x 10^{-5} per ppb at standard temperature and pressure), but a reference concentration was not established.[127]

Estimating cancer risk

The Agency for Toxic Substances and Disease Registry (ATSDR) provides the following equation for calculating cancer risk:

$$CR = (AAC \times IUR) \times \left(\frac{ED}{LY}\right) \quad (3)$$

Where CR is the cancer risk, AAC is the exposure factor-adjusted air concentration, IUR is the inhalation unit risk, ED is the age-specific exposure duration, and LY is the lifetime in years. We analyzed the worst case-scenario where the exposure duration is equal to the lifetime, and the exposure factor is one (i.e., continuous exposure). Under these assumptions, the cancer risk is

estimated as the product of the air concentration and the IUR. A more accurate estimate requires information about individual exposure patterns, but given that formaldehyde concentrations are typically higher indoors,[128] assuming continuous exposure to ambient concentrations as a baseline is reasonable.

Results and discussion

Sampling success rates

A total of 55 samples were collected by participants and sent to the laboratory for analysis, including eight field blanks. The analytical method quantified several carbonyls on each sample, but we only present formaldehyde results given that we did not implement all the required protocols listed in EPA's method TO-11A for the reliable detection of other aldehydes and ketones.[129] Sensor data were analyzed to assess sample quality. Quality assurance criteria were applied sequentially as described hereafter. Out of the 47 samples collected (not counting blanks), twelve (26%) had missing or invalid log files because the host did not start the sampler correctly or, in the case of one device, because the memory card was lost. Five samples (11%) were considered invalid because the runtime was short (< 12 hours), which was caused by low battery levels at start-up in most cases, and by a firmware issue in one case. Two additional samples (4%) were invalidated because a blockage was detected in the sampling path (see Figures C2 and C3). The application of these quality assurance criteria resulted in 28 valid samples (60%).

Approximately 85% of the sample attrition was caused by human errors, underscoring the importance of robust training and follow-up communication for community-based sampling. Nonetheless, the significant number of issues with starting samples and charging shows that

technological improvements (e.g., simplify operation, implement remote control/monitoring) are also required to make the in-field protocols and operation more dependable for community-based monitoring. The remaining 15% of the sample attrition was caused by issues with the samplers, which represent a small fraction of the total samples (6%). Improvements to hardware and software reliability are desirable, but they would probably have a minor effect on the overall success rates.

Four samples (9%) had formaldehyde peak areas below the limit of detection (LOD) calculated using the field blanks (see detailed LOD calculations in Appendix C section 2.2). The mean sample duration was 49 hours, and the median was 59 hours (IQR = 30 h). The mean sample volume was 91 L, and the median was 93 L (IQR = 33 L). For both the mean and median sample volume, the LOD was equivalent to a formaldehyde concentration in air of approximately 0.33 ppb. All the samples below the LOD had shorter than average durations (13 – 26 h), so our sampling methods are adequate for multi-day samples, but not for single-day or shorter samples.

More effort is needed to ensure a minimum sample duration of ~48 hours. The shortest samples were caused by insufficient recharging of the batteries between samples. Modifying the charging procedures to make them easier to follow, increasing the charging speed, having a spare battery that could be charged while the other battery is in-use, or adding solar charging are possible changes that could have a positive effect on sample duration consistency. Another effective intervention could be to communicate that starting a sample late with a fully charged battery is better than starting the sample in a timely manner with a partially charged battery. A small number of samples (~5) that started with fully charged batteries had durations in the 24-h to 48-h range. This was caused by an improperly installed auxiliary board that increased power draw. The removal and installation of the auxiliary board was covered in detail during the in-person

training and in the user manual, but evidence shows that engineering changes are necessary to effectively avoid these issues arising in the field.

Formaldehyde ambient concentrations and cancer risks

Table 4: Formaldehyde concentrations, cancer risks, and sample size statistics classified by super neighborhood. Time-weighted averages are based on all the samples collected in each super neighborhood and may not cover every week of the sampling campaign.

Super neighborhood	*Time-weighted average [ppb]*	*Range [ppb]*	*Cancer risk*	*Number of samples*	*Total sampling time [h]*
Clinton Park	3.01	0.95 – 5.98	4.07×10^{-5}	7	168
Pleasantville Area	2.58	1.30 – 5.35	3.49×10^{-5}	5	282
Galena Park	2.16	1.27 – 3.53	2.92×10^{-5}	4	251
Greater Fifth Ward	4.22	2.41 – 7.30	5.70×10^{-5}	3	108
Greater Heights	2.10	1.29 – 3.14	2.84×10^{-5}	6	292
Greater Third Ward	1.72	1.59 – 1.88	2.32×10^{-5}	3	202
Sunnyside	2.22	2.22 – 2.22	3.00×10^{-5}	1	64
Trinity/Houston Gardens	2.36	1.32 – 4.19	3.19×10^{-5}	2	102

We quantified formaldehyde concentrations for 24 validated samples with peak areas above the LOD. In addition to the samples we collected during our community-based campaign, we retrieved data for 7 samples collected by TCEQ during the same period at their station in Clinton Park, a super neighborhood adjacent to Pleasantville and Galena Park. Concentration statistics are summarized in Table 4. Weekly time-time weighted average concentrations are presented in Figure 11 classified by super neighborhood. Most samples (58% of community-collected and 43% of TCEQ-collected samples) were lower than TCEQ's chronic noncarcinogenic nonlinear ESL (i.e., the lowest reference value for ambient formaldehyde concentrations in Texas). The rest of the samples, except for one TCEQ sample, were higher than TCEQ's noncarcinogenic nonlinear ESL, but lower than EPA's IRIS RfC. One of the TCEQ samples was slightly higher than EPA's IRIS RfC.

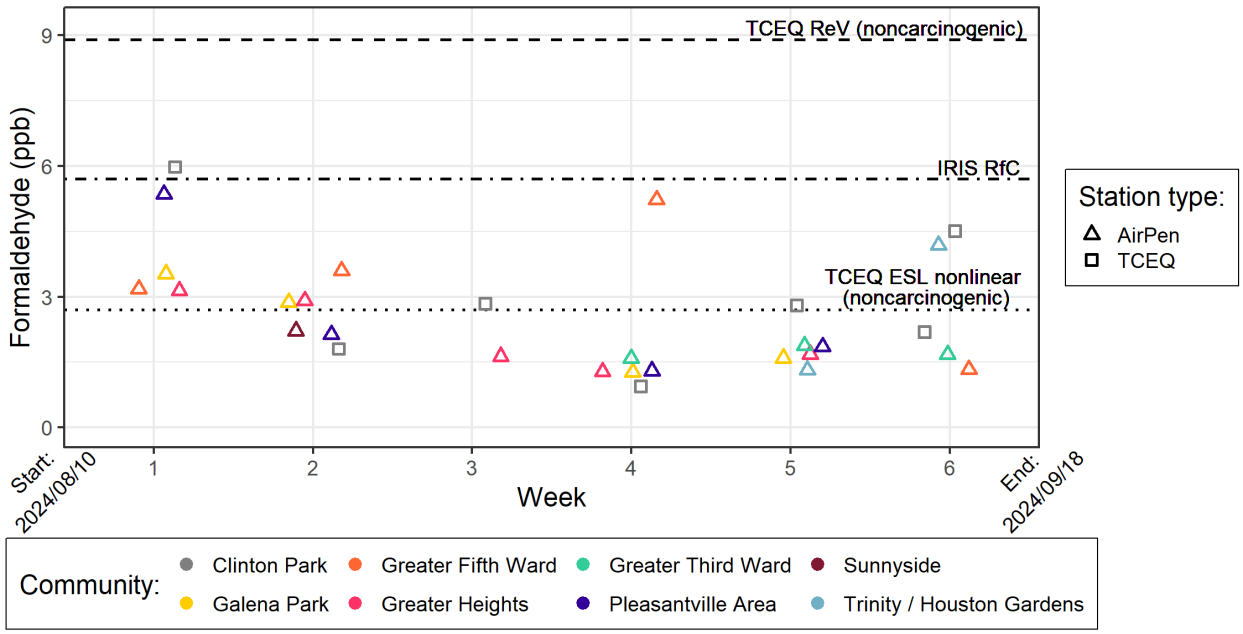

Figure 11: Weekly time-weighted average formaldehyde mixing ratios (i.e., concentrations) classified by community (i.e., super neighborhood) and monitoring station type (i.e., the AirPen-based monitors we deployed vs. the regulatory monitoring station run by TCEQ). Random jitter about the x-axis was added to individual points to separate them and facilitate visualization. Exact start times are not reflected on the position along the x-axis. The TCEQ ReV (noncarcinogenic) is the long-term (chronic) reference value used for the evaluation of air monitoring data in Texas. The TCEQ ESL nonlinear (noncarcinogenic) is the long-term (chronic) effects screening level used for air permit reviews in Texas. The IRIS RfC is the inhalation reference concentration identified by EPA's 2024 toxicology review.

Based on the concentrations we measured, special screening efforts are likely not required for noncarcinogenic effects. For instance, TCEQ only applies the noncarcinogenic nonlinear ESL for air permitting and not for regular air monitoring, so its exceedance is not an infringement, but it indicates that a health concern may exist necessitating additional monitoring. Furthermore, the EPA IRIS health assessment considers that noncarcinogenic effects are not expected for the concentrations presented in this work.[127] Nevertheless, the carcinogenic effects of formaldehyde described in the IRIS assessment warrant an analysis of the increased cancer risk at any measurable concentration. The cancer risks, estimated as the product of the time-weighted average concentrations for each super neighborhood and the IUR given by the IRIS toxicological review, ranged from 2.32×10^{-5} to 5.70×10^{-5} (see Table 4) for a lifetime exposure. According to

ATSDR, other factors (e.g., known carcinogenicity in humans, exposure of vulnerable populations) must be considered to determine if cancer risks in the 1×10^{-6} to 1×10^{-4} range are a public health concern.[130]

Evaluating cancer risk involves numerous assumptions about exposure duration, concentration variability, and human sensitivity and is, therefore, subject to major uncertainty. The cancer risks we estimated are based on a few weeks of concentration data and correspond only to the exposure to formaldehyde in ambient air. Exposure to other formaldehyde sources and other carcinogens could increase cancer risks beyond the 1×10^{-4} threshold. Several of the super neighborhoods included in this study have statistically higher than normal cancer rates and have known sources of carcinogens.[122,131,132] Crucially, Fifth Ward has already been identified as a cancer cluster by the Houston Health Department due to creosote contamination; distinct from that determination by the Houston Health Department, in this study we estimated the formaldehyde-induced cancer risk in Fifth Ward to be the highest among our monitoring sites. The second and third highest cancer risks were estimated for Clinton Park and Pleasantville. These three communities are located along the Houston Ship Channel. Our findings reinforce evidence for public health concern among communities along the channel.

Strengths and key findings

This study demonstrates that regulatory monitoring of HAPs can be complemented by community-based monitoring using lower-cost instruments to improve spatial and temporal coverage. In this first deployment, we obtained HAP measurements for seven super neighborhoods, whereas TCEQ has stations that monitor HAPs in two super neighborhoods in Houston. Scaling up the number of stations would be relatively inexpensive using our platform (~$2000 per sampler). The low instrument cost and self-supported functionality (i.e., no grid

power required) would also allow us to monitor under conditions that force TCEQ (and probably other agencies) to shut down regulatory monitors, filling the spatial and temporal gaps, and complementing the government-led efforts. Even as we adhere to mandated sampling media and analytical methods, shifting toward compact, budget-friendly samplers and accessible field protocols has the added benefit of empowering local groups to sample during key events (i.e., for disaster response or when other monitors are offline) in their locality.

Comprehensive on-board sensor diagnostics enabled thorough sample quality assessment. The approach to sample collection, analysis, and quality assurance and control presented in this work ensured that faults were identified and provided the data necessary to determine the cause of those faults, both of which were essential to obtaining trustworthy results. Samples that passed the quality assurance criteria had concentrations consistent with previously reported data.[133] Moreover, the monitors in Pleasantville and Galena Park always measured concentrations within 1 ppb of the values reported by the TCEQ station in the contiguous super neighborhood of Clinton Park. AirPen sampling performance and the analytical method have been previously evaluated and are out of the scope of this work,[19,129] but the qualitative comparison with the status quo data shows that our results are tenable.

Limitations and future work

The findings of this study also show the need for technological and procedural improvements to increase sample success rates under field deployed conditions. Sampler launch was a major point of failure, so the user interface must be improved. Battery recharging can also be modified to rely less on volunteers' abilities to disassemble the sampler and plug the battery to a wall charger. Additionally, having remote control and monitoring of the samplers during deployment and sampling would enable the identification of issues in real time and increase the likelihood of

their prompt resolution, further increasing overall sample success. Here, we deemed a sample successful when it produced a valid formaldehyde result. However, the utility of this approach could be greatly enhanced by extending the same cartridge-based analysis to additional HAPs, as outlined in EPA Method TO-11A. Long-term campaigns that adopt the methodology improvements outlined above (i.e., simplified sampler launch, streamlined battery-swap procedures, remote fault detection, and the ability to quantify multiple carbonyls per cartridge) would greatly strengthen community-based HAP risk assessments and refine cancer-risk estimations.

CHAPTER 5. CONCLUSIONS

This work presents key steps required to develop a monitoring platform with the potential to fill in the gaps left by reference monitoring networks (for ambient air), and traditional industrial hygiene methods (for personal exposure assessment). Throughout the chapters, we cover mechanical, electrical, and software implementations, testing and validation of individual components (i.e., sensors), comprehensive field demonstrations of the platform, and analytical frameworks for the quality assurance of sampling results. Equally important are the real-world findings of our deployments that show the need for higher sampling throughput and coverage. The development of the AirPen platform predates this dissertation, and its evolution will continue beyond it, demonstrating that the ultimate goal of delivering an accessible and reliable alternative to established technology is not as simple as integrating an array of sensors and logging raw data.

The novel and established methodologies described here were selected to streamline the collection of defensible data. The use of new microelectronics lowered instrument cost and size significantly (relative to reference grade monitors), deployments were simplified through technological innovation (e.g., automation, pre-configuration, phone-based interfaces), and on-board diagnostics made quality assurance and control manageable even for relatively large deployments (i.e., > 50 concurrent samples). We also reduced the expert support needed in the field and, by engaging volunteers and community partners, transformed them into integral contributors to the sampling scheme. These advancements allowed us to collect unprecedented amounts of data for each population in our studies (i.e., a 5 – 10 increase in sampling throughput relative to the *status quo*), considering the scope of our projects and the available resources.

The platform has evolved over the years to incorporate feedback from our field experiences and to implement new functionality (e.g., new sensors, different sampling media) as a response to the interests of research partners. However, the underlying AirPen technology has persisted due to its distinctive ability to collect particulate and gas samples, as well as sensor data, simultaneously. Although sample preparation and analysis require different methodologies depending on the analytes in question, and even the instrument is slightly modified for each application (e.g., housings, batteries, and adapters can be replaced), many of the data analysis and validation frameworks remain unchanged or require minor modifications. Therefore, different applications can be conveniently accommodated by selecting appropriate components and sampling media. Monitors can be stationary, portable, or wearable. PM, CO_2, light, noise, and various other sensors can be installed as needed. $PM_{2.5}$ and inhalable size-selective inlets are available for filter-based sampling, and gas samples can be collected into sorbent tubes of different types or DNPH cartridges. Supplementary batteries can be selected to fit the expected sampling duration. This modularity makes the AirPen a highly adaptable platform.

For the AirPen to become a break-through technology, several practical issues and limitations identified in this work must still be addressed. While sample attrition was non-trivial in both field demonstrations, the lessons learned point directly to actionable design refinements. For the facility-wide exposure assessment covered in chapter 3, most of the samples were lost due to sample media damage during shipping. This was addressed by switching from glass tubes to plastic cartridges for carbonyl sampling. For the stationary monitoring campaign covered in chapter 4, participants struggled to start samples and to recharge the batteries properly, evidencing the need for usability improvements. These usability changes have been implemented or will be implemented in the short-term for next projects. Cellular connectivity has been

implemented, allowing users to start a sample through a web-based interface by scanning two QR codes, and researchers can monitor the status of each sampler remotely. New solar charging capabilities enable uninterrupted monitoring under fair weather conditions, until memory becomes a bottleneck (> a week per sample). Instrument reliability has been good, but minor software bugs and a small number of leaks and blockages in the sampling train have been found. Hardware and software improvements were implemented after the demonstration deployments and continue to be explored.

In the mid-term, as the platform becomes more robust, the focus can shift to expanding monitoring capabilities. For example, the same cartridges used for formaldehyde sampling could be used for other aldehydes and ketones, including several hazardous air pollutants, by modifying the analytical method slightly.[129] Better sensors could improve the real-time monitoring capabilities (e.g., a photoionization detector could replace the metal oxide sensor for VOC monitoring). Finally, the software could be modified to enable active flow control through the gas tube, as an alternative to active flow control through the particle filter (one at a time, not simultaneously). Each of the improvements listed above is relatively straightforward; taken together, they constitute an exciting agenda for future research and commercialization.

Nonetheless, the key findings of our real-world studies (e.g., major variability in exposure within populations, gaps in monitoring during high-risk events, measurable health concerns) underscore the importance of continuing the long-term development of this technology and expanding its reach. Large-scale longitudinal deployments, covering populations typically underrepresented by current exposure assessment efforts, could advance the state of knowledge in various fields (e.g., environmental justice, exposomics, epidemiology). Collectively, these findings chart a clear path

toward and position the AirPen to make a substantive impact on exposure science and public health.

REFERENCES

(1) Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2021*; Seattle, WA, 2024. https://vizhub.healthdata.org/gbd-results/.

(2) Kim, K. H.; Kabir, E.; Kabir, S. A Review on the Human Health Impact of Airborne Particulate Matter. *Environ. Int.* **2015**, *74*, 136–143. https://doi.org/10.1016/j.envint.2014.10.005.

(3) Forouzanfar, M. H.; Alexander, L.; Bachman, V. F.; Biryukov, S.; Brauer, M.; Casey, D.; Coates, M. M.; Delwiche, K.; Estep, K.; Frostad, J. J.; Astha, K. C.; Kyu, H. H.; Moradi-Lakeh, M.; Ng, M.; Slepak, E.; Thomas, B. A.; Wagner, J.; Achoki, T.; Atkinson, C.; Barber, R. M.; Cooperrider, K.; Dandona, L.; Dicker, D.; Flaxman, A. D.; Fleming, T. D.; Foreman, K. J.; Gakidou, E.; Hay, S. I.; Heuton, K. R.; Iannarone, M. L.; Ku, T.; Larson, H. J.; Lim, S. S.; Lopez, A. D.; Lozano, R.; MacIntyre, M. F.; Margono, C.; McLain, A.; Mokdad, A. H.; Mullany, E. C.; Murray, C. J. L.; Naghavi, M.; Nguyen, G.; Pain, A. W.; Richardson, L.; Robinson, M.; Sandar, L.; Stephens, N.; Temesgen, A. M.; Thomson, B.; Vos, T.; Wan, X.; Wang, H.; Wurtz, B.; Ebel, B. E.; Ellenbogen, R. G.; Wright, J. L.; Alfonso-Cristancho, R.; Anderson, B. O.; Jensen, P. N.; Quistberg, D. A.; Riederer, A.; Vavilala, M. S.; Zunt, J. R.; Anderson, H. R.; Pourmalek, F.; Gotay, C. C.; Burnett, R.; Shin, H. H.; Weichenthal, S.; Cohen, A.; Knudsen, A.; Aasvang, G.; Kinge, J. M.; Skirbekk, V.; Vollset, S.; Abbafati, C.; Abbasoglu Ozgoren, A.; Çavlin, A.; Kucuk Bicer, B.; Abd-Allah, F.; Abera, S. F.; Melaku, Y. A.; Aboyans, V.; Abraham, B.; Puthenpurakal Abraham, J.; Abraham, J. P.; Thorne-Lyman, A. L.; Ding, E. L.; Fahimi, S.; Khatibzadeh, S.; Wagner, G. R.; Bukhman, G.; Campos-Nonato, I. R.; Feigl, A. B.; Salomon, J. A.;

Benzian, H.; Abubakar, I.; Abu-Rmeileh, N. M. E.; Aburto, T. C.; Avila, M. A.; Barquera, S.; Barrientos-Gutierrez, T.; Campuzano, J. C.; Cantoral, A. J.; Contreras, A. G.; Cuevas-Nasu, L.; De, V.; García-Guerra, F. A.; Gomez Dantes, H.; Gonzalez de Cosio, T.; González-Castell, D.; Heredia-Pi, I. B.; Hernandez, L.; Jauregui, A.; Medina, C.; Mejia-Rodriguez, F.; Montañez Hernandez, J. C.; Pedraza, L. S.; Pedroza, A.; Quezada, A. D.; Salvo, D.; Sanchez, L. M.; Sánchez-Pimienta, T. G.; Servan-Mori, E. E.; Shamah Levy, T.; Téllez Rojo, M. M.; Villalpando, S.; Adelekan, A.; Adofo, K.; Adou, A. K.; Adsuar, J. C.; Fra Paleo, U.; Afshin, A.; Micha, R.; Mozaffarian, D.; Shahraz, S.; Shangguan, S.; Singh, G. M.; Agardh, E. E.; Al Khabouri, M. J.; Al Lami, F. H.; Alam, S.; Naheed, A.; Alasfoor, D.; Albittar, M. I.; Alegretti, M. A.; Cavalleri, F.; Aleman, A. V.; Colistro, V.; Alemu, Z. A.; Alhabib, S.; Chen, Z.; Gething, P.; Ali, R.; Bennett, D. A.; Briggs, A. D. M.; Rahimi, K.; Scarborough, P.; Simard, E. P.; Ali, M. K.; Argeseanu Cunningham, S.; Liu, Y.; Narayan, K. M. V.; Omer, S. B.; Alla, F.; Guillemin, F.; Allebeck, P.; Roy, N.; Kivipelto, M.; Weiderpass, E.; Fereshtehnejad, S.; Havmoeller, R.; Sindi, S.; Allen, P. J.; Alsharif, U.; Endres, M.; Nolte, S.; Papachristou, C.; Alvarez, E.; Alvis-Guzman, N.; Paternina Caicedo, A. J.; Amankwaa, A. A.; Amare, A. T.; Hoek, H. W.; Gansevoort, R. T.; Yenesew, M.; Ameh, E. A.; Ameli, O.; Amini, H.; Tanner, M.; Ammar, W.; Harb, H. L.; Antonio, C. A. T.; Faraon, E. A.; Panelo, C. A.; Anwari, P.; Arnlöv, J.; Larsson, A.; Arsic Arsenijevic, V. S.; Artaman, A.; Asghar, R. J.; Assadi, R.; Atkins, L. S.; Awuah, B.; Laryea, D. O.; Badawi, A.; Bahit, M. C.; Bakfalouni, T.; Balakrishnan, K.; Balalla, S.; Feigin, V. L.; Te Ao, B. J.; Balu, R.; Dandona, R.; Goenka, S.; Kumar, G.; Murthy, K. S.; Reddy, K.; Banerjee, A.; Barker-Collo, S. L.; del Pozo-Cruz, B.; Barregard, L.; Barrero, L. H.; Basto-Abreu, A. C.; Batis Ruvalcaba, C.; de Castro, E. F.; Lopez, N.; Texcalac, J.

L.; Basu, A.; Gaffikin, L.; Basu, S.; Basulaiman, M. O.; Memish, Z. A.; Beardsley, J.; Bedi, N.; Bekele, T.; Bell, M. L.; Huang, J. J.; Benjet, C.; Borges, G.; Gutiérrez, R. A.; Orozco, R.; Trasande, L.; Hagan, H.; Bernabé, E.; Wolfe, C. D. A.; Beyene, T. J.; Bhala, N.; Derrett, S.; Bhalla, A.; Jha, V.; Bhutta, Z. A.; Nisar, M. I.; Bikbov, B.; Bin Abdulhak, A. A.; Vijayakumar, L.; Chiang, P. P.; Blore, J. D.; Brooks, P. M.; Lakshmana Balaji, A.; Colquhoun, S. M.; Weintraub, R. G.; Blyth, F. M.; Meretoja, A.; Bohensky, M. A.; Bora Basara, B.; Yentür, G. K.; Kose, M. R.; Pekericli, A.; Uzun, S. B.; Bornstein, N. M.; Bose, D.; Boufous, S.; Degenhardt, L.; Bourne, R. R.; Brainin, M.; Brazinova, A.; Majdan, M.; Breitborde, N. J.; Schöttker, B.; Brenner, H.; Broday, D. M.; Lunevicius, R.; Bruce, N. G.; Dherani, M. K.; Pope, D.; Brugha, T. S.; Brunekreef, B.; Kromhout, H.; Buchbinder, R.; Gabbe, B.; Gibney, K. B.; Thrift, A. G.; Bui, L. N.; Nguyen, N. T.; Bulloch, A. G.; Patten, S. B.; Tonelli, M.; Wang, J.; Burch, M.; Burney, P. G. J.; Jarvis, D. L.; Rodriguez, A.; Rushton, L.; Soljak, M.; Williams, T. N.; Caravanos, J.; Nash, D.; Cárdenas, R.; Cardis, E.; Nieuwenhuijsen, M. J.; Rojas-Rueda, D.; Carpenter, D. O.; Leung, R.; Caso, V.; Castañeda-Orjuela, C. A.; Castro, R. E.; Catalá-López, F.; Chadha, V. K.; Chang, J.; Scott, J. G.; Hoy, D. G.; Knibbs, L. D.; Charlson, F. J.; Erskine, H. E.; Ferrari, A. J.; Gouda, H. N.; Veerman, L. J.; Whiteford, H. A.; Chen, W.; Zou, X.; Chen, H.; London, S. J.; Jiang, Y.; Takahashi, K.; Chimed-Ochir, O.; Chowdhury, R.; Powles, J.; Christophi, C. A.; Chuang, T.; Chugh, S. S.; Cirillo, M.; Claßen, T. K. D.; Kraemer, A.; Tobollik, M.; Colomar, M.; Cooper, C.; Cooper, L. T.; Coresh, J.; Matsushita, K.; Tran, B. X.; Courville, K. J.; Criqui, M. H.; Stein, M. B.; Damsere-Derry, J.; Danawi, H.; Refaat, A. H.; Dargan, P. I.; Davis, A.; Fay, D. F. J.; Schmidt, J. C.; Davitoiu, D. V.; Dayama, A.; DeLeo, D.; de Lima, G.; Machado, V. M. P.; Nogueira, J. R.; Teixeira, C. M.; Dellavalle,

R. P.; Deribe, K.; Mekonnen, W.; Des Jarlais, D. C.; Dessalegn, M.; deVeber, G. A.; Lindsay, M. P.; Hu, H.; Devries, K. M.; McKee, M.; Pearce, N.; Stöckl, H.; Tillmann, T.; Watts, C. H.; Dharmaratne, S. D.; Dokova, K.; Dorsey, E. R.; Driscoll, T. R.; Marks, G. B.; Leigh, J.; Duan, L.; Li, Y.; Liu, S.; Ma, J.; Wang, L.; Ye, P.; Zhou, M.; Liang, X.; Durrani, A. M.; Elshrek, Y. M.; Ermakov, S. P.; Soshnikov, S.; Eshrati, B.; Farzadfar, F.; Esteghamati, A.; Hafezi-Nejad, N.; Sheikhbahaei, S.; Sepanlou, S. G.; Heydarpour, P.; Sahraian, M.; Rahimi-Movaghar, V.; Ferri, C. P.; Foigt, N.; Franklin, R. C.; Gamkrelidze, A.; Khonelidze, I.; Sturua, L.; Gankpé, F. G.; Gasana, E.; Sabin, N.; Geleijnse, J. M.; Gessner, B. D.; Gillum, R. F.; Ginawi, I. A. M.; Giroud, M.; Giussani, G.; Goginashvili, K.; Gona, P.; Goto, A.; Guerrant, R. L.; Terkawi, A. S.; Gugnani, H. C.; Gunnell, D.; Gupta, R.; Hagstromer, M.; Halasa, Y. A.; Idrisov, B. T.; Hamadeh, R. R.; Hammami, M.; Hankey, G. J.; Hao, Y.; Zheng, Y.; Haregu, T.; van de Vijver, S.; Haro, J.; Hedayati, M. T.; Hijar, M.; Hoffman, H. J.; Mensah, G. A.; Sampson, U. K.; Hornberger, J. C.; Hosgood, H.; Hsairi, M.; Hu, G.; Huang, C.; Hubbell, B. J.; Huiart, L.; Racapé, L.; Husseini, A.; Iburg, K. M.; Ikeda, N.; Innos, K.; Inoue, M.; Kawakami, N.; Shibuya, K.; Islami, F.; Ismayilova, S.; Jacobsen, K. H.; Jansen, H. A.; Jassal, S. K.; Jayaraman, S.; Jeemon, P.; Prabhakaran, D.; Jiang, F.; Jiang, G.; Phillips, M. R.; Jonas, J. B.; Juel, K.; She, J.; Kan, H.; Kany Roseline, S. S.; Karam, N. E.; Karch, A.; Karema, C. K.; Karthikeyan, G.; Paul, V. K.; Satpathy, M.; Tandon, N.; Kaul, A.; Kazi, D. S.; Kemp, A. H.; Lotufo, P. A.; Polanczyk, G. V.; Santos, I. S.; Kengne, A. P.; Matzopoulos, R.; Parry, C. D.; Sliwa, K.; Mayosi, B. M.; Stein, D. J.; Keren, A.; Khader, Y. S.; Ali Hassan Khalifa, S. E.; Khan, E. A.; Khang, Y.; Kieling, C.; Kim, D.; Kim, S.; Kim, Y.; Kimokoti, R. W.; Kinfu, Y.; Kissela, B. M.; Kokubo, Y.; Kosen, S.; Warouw, T. S.; Kravchenko, M.;

Varakin, Y. Y.; Krishnaswami, S.; Kuate Defo, B.; Kuipers, E. J.; Polinder, S.; Kulkarni, C.; Kulkarni, V. S.; Kwan, G. F.; Lai, T.; Lalloo, R.; Lallukka, T.; Shiri, R.; Lam, H.; Lan, Q.; Lansingh, V. C.; Lavados, P. M.; Lawrynowicz, A. E.; Leasher, J. L.; Lee, J.; Yoon, S.; Levi, M.; Liang, J.; Wang, Y.; Zhu, J.; Lipshultz, S. E.; Lloyd, B. K.; Room, R.; Logroscino, G.; Lortet-Tieulent, J.; Ma, S.; Phua, H. P.; Magis-Rodriguez, C.; Mahdi, A. A.; Malekzadeh, R.; Mangalam, S.; Mapoma, C. C.; Masiye, F.; Marape, M.; Marcenes, W.; Meaney, P. A.; Margolis, D. J.; Silberberg, D. H.; Martin, R. V.; Marzan, M. B.; Mashal, M. T.; Mason-Jones, A. J.; Mazorodze, T. T.; McKay, A. C.; Mehndiratta, M.; Meltzer, M.; Mendoza, W.; Apolinary Mhimbira, F.; Miller, T. R.; Mills, E. J.; Mishra, S.; Mohamed Ibrahim, N.; Mohammad, K. A.; Mola, G. L.; Monasta, L.; Montico, M.; Ronfani, L.; Moore, A. R.; Morawska, L.; Norman, R. E.; Mori, R.; Tsilimbaris, M.; Moschandreas, J.; Moturi, W. N.; Werdecker, A.; Mueller, U. O.; Westerman, R.; Mukaigawara, M.; Nahas, Z.; Naidoo, K. S.; Naldi, L.; Nand, D.; Nangia, V.; Neal, B.; Nejjari, C.; Neupane, S. P.; Newton, C. R.; Ngalesoni, F. N.; Ngirabega, J. D.; Nolla, J. M.; Vollset, S. E.; Norheim, O. F.; Norrving, B.; Nyakarahuka, L.; Oh, I.; Ohkubo, T.; Olusanya, B. O.; Opio, J. N.; Pagcatipunan, R. S.; Pandian, J. D.; Park, E.; Seedat, S.; Pavlin, B. I.; Pejin Stokic, L.; Pereira, D. M.; Perez-Padilla, R.; Perez-Ruiz, F.; Perico, N.; Remuzzi, G.; Trillini, M.; Perry, S. A. L.; Pervaiz, A.; Pesudovs, K.; Peterson, C. B.; Petzold, M.; Plass, D.; Poenaru, D.; Pond, C. D.; Pope, C.; Popova, S.; Rehm, J.; Prasad, N. M.; Qato, D. M.; Rafay, A.; Rana, S. M.; Ur Rahman, S.; Raju, M.; Rakovac, I.; Rao, M.; Razavi, H.; Ribeiro, A. L.; Velasquez-Melendez, G.; Riccio, P. M.; Sposato, L. A.; Roca, A.; Romieu, I.; Straif, K.; Ruhago, G. M.; Sunguya, B. F.; Sacco, R. L.; Saha, S.; Sahathevan, R.; Sanabria, J. R.; Sanchez-Riera, L.; Sapkota, A.; Saunders, J. E.; Soneji,

S.; Sawhney, M.; Saylan, M. I.; Schneider, I. J. C.; Schwebel, D. C.; Singh, J. A.; Serdar, B.; Shaddick, G.; Shinohara, Y.; Shishani, K.; Shiue, I.; Sigfusdottir, I. D.; Singh, A.; Søreide, K.; Sreeramareddy, C. T.; Stapelberg, N. J. C.; Stathopoulou, V.; Steckling, N.; Stroumpoulis, K.; Swaminathan, S.; Swaroop, M.; Yano, Y.; Sykes, B. L.; Tabb, K. M.; Talongwa, R. T.; Tanne, D.; Tavakkoli, M.; Thackway, S. V.; Thurston, G. D.; Topouzis, F.; Towbin, J. A.; Toyoshima, H.; Traebert, J.; Trujillo, U.; Tsala Dimbuene, Z.; Tuzcu, E.; Uchendu, U. S.; Ukwaja, K. N.; Van Dingenen, R.; van Gool, C. H.; van Os, J.; Vasankari, T. J.; Vasconcelos, A. N.; Violante, F. S.; Victorovich Vlassov, V.; Waller, S. G.; Wallin, M. T.; Wang, W.; Wessells, K.; Wilkinson, J. D.; Williams, H. C.; Woldeyohannes, S. M.; Wong, J. Q.; Woolf, A. D.; Xu, G.; Yan, L. L.; Yang, G.; Yip, P.; Yonemoto, N.; Younis, M. Z.; Younoussi, Z.; Yu, C.; Zaki, M. E.; Zhao, Y.; Zhu, S. Global, Regional, and National Comparative Risk Assessment of 79 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks in 188 Countries, 1990-2013: A Systematic Analysis for the Global Burden of Disease Study 2013. *Lancet* **2015**, *386* (10010), 2287–2323. https://doi.org/10.1016/S0140-6736(15)00128-2.

(4) Brauer, M.; Freedman, G.; Frostad, J.; Van Donkelaar, A.; Martin, R. V.; Dentener, F.; Dingenen, R. Van; Estep, K.; Amini, H.; Apte, J. S.; Balakrishnan, K.; Barregard, L.; Broday, D.; Feigin, V.; Ghosh, S.; Hopke, P. K.; Knibbs, L. D.; Kokubo, Y.; Liu, Y.; Ma, S.; Morawska, L.; Sangrador, J. L. T.; Shaddick, G.; Anderson, H. R.; Vos, T.; Forouzanfar, M. H.; Burnett, R. T.; Cohen, A. Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environ. Sci. Technol.* **2016**, *50* (1), 79–88. https://doi.org/10.1021/acs.est.5b03709.

(5) Krishnamoorthy, K.; Mathew, T. Inferences on the Means of Lognormal Distributions Using Generalized P-Values and Generalized Confidence Intervals. *J. Stat. Plan. Inference* **2003**, *115* (1), 103–121. https://doi.org/10.1016/S0378-3758(02)00153-2.

(6) Rappaport, S. M. ASSESSMENT OF LONG-TERM EXPOSURES TO TOXIC SUBSTANCES IN AIR. *Ann. Occup. Hyg.* **1991**, *35* (1), 61–122. https://doi.org/10.1093/annhyg/35.1.61.

(7) Lyles, R. H.; Kupper, L. L.; Rappaport, S. M. A LOGNORMAL DISTRIBUTION-BASED EXPOSURE ASSESSMENT METHOD FOR UNBALANCED DATA. *Ann. Occup. Hyg.* **1997**, *41* (1), 63–76. https://doi.org/10.1093/annhyg/41.1.63.

(8) Waters, M. A.; Selvin, S.; Rappaport, S. M. A MEASURE OF GOODNESS-OF-FIT FOR THE LOGNORMAL MODEL APPLIED TO OCCUPATIONAL EXPOSURES. *Am. Ind. Hyg. Assoc. J.* **1991**, *52* (11), 493–502. https://doi.org/10.1080/15298669191365108.

(9) Seixas, N. S.; Robins, T. G.; Moulton, L. H. The Use of Geometric and Arithmetic Mean Exposures in Occupational Epidemiology. *Am. J. Ind. Med.* **1988**, *14* (4), 465–477. https://doi.org/https://doi.org/10.1002/ajim.4700140410.

(10) Hajat, A.; Hsia, C.; O'Neill, M. S. Socioeconomic Disparities and Air Pollution Exposure: A Global Review. *Curr. Environ. Heal. reports* **2015**, *2* (4), 440–450. https://doi.org/10.1007/s40572-015-0069-5.

(11) Nguyen, N. P.; Marshall, J. D. Impact, Efficiency, Inequality, and Injustice of Urban Air Pollution: Variability by Emission Location. *Environ. Res. Lett.* **2018**, *13* (2). https://doi.org/10.1088/1748-9326/aa9cb5.

(12) Ródenas García, M.; Spinazzé, A.; Branco, P. T. B. S.; Borghi, F.; Villena, G.; Cattaneo, A.; Di Gilio, A.; Mihucz, V. G.; Gómez Álvarez, E.; Lopes, S. I.; Bergmans, B.; Orłowski, C.; Karatzas, K.; Marques, G.; Saffell, J.; Sousa, S. I. V. Review of Low-Cost Sensors for Indoor Air Quality: Features and Applications. *Appl. Spectrosc. Rev.* **2022**, *57* (9–10), 747–779. https://doi.org/10.1080/05704928.2022.2085734.

(13) Kang, Y.; Aye, L.; Ngo, T. D.; Zhou, J. Performance Evaluation of Low-Cost Air Quality Sensors: A Review. *Sci. Total Environ.* **2022**, *818*, 151769. https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.151769.

(14) Sayahi, T.; Butterfield, A.; Kelly, K. E. Long-Term Field Evaluation of the Plantower PMS Low-Cost Particulate Matter Sensors. *Environ. Pollut.* **2019**, *245*, 932–940. https://doi.org/10.1016/j.envpol.2018.11.065.

(15) Bulot, F. M. J.; Johnston, S. J.; Basford, P. J.; Easton, N. H. C.; Apetroaie-Cristea, M.; Foster, G. L.; Morris, A. K. R.; Cox, S. J.; Loxham, M. Long-Term Field Comparison of Multiple Low-Cost Particulate Matter Sensors in an Outdoor Urban Environment. *Sci. Rep.* **2019**, *9* (1), 1–13. https://doi.org/10.1038/s41598-019-43716-3.

(16) Papaconstantinou, R.; Demosthenous, M.; Bezantakos, S.; Hadjigeorgiou, N.; Costi, M.; Stylianou, M.; Symeou, E.; Savvides, C.; Biskos, G. Field Evaluation of Low-Cost Electrochemical Air Quality Gas Sensors under Extreme Temperature and Relative Humidity Conditions. *Atmos. Meas. Tech.* **2023**, *16* (12), 3313–3329. https://doi.org/10.5194/amt-16-3313-2023.

(17) Tryner, J.; L'Orange, C.; Mehaffy, J.; Miller-Lionberg, D.; Hofstetter, J. C.; Wilson, A.; Volckens, J. Laboratory Evaluation of Low-Cost PurpleAir PM Monitors and in-Field

Correction Using Co-Located Portable Filter Samplers. *Atmos. Environ.* **2020**, *220* (June 2019), 117067. https://doi.org/10.1016/j.atmosenv.2019.117067.

(18) Goletto, V.; Mialon, G.; Faivre, T.; Wang, Y.; Lesieur, I.; Petigny, N.; Vijapurapu, S. S. Formaldehyde and Total VOC (TVOC) Commercial Low-Cost Monitoring Devices: From an Evaluation in Controlled Conditions to a Use Case Application in a Real Building. *Chemosensors* **2020**, *8* (1). https://doi.org/10.3390/chemosensors8010008.

(19) Tryner, J.; Quinn, C.; Molina Rueda, E.; Andales, M. J.; L'Orange, C.; Mehaffy, J.; Carter, E.; Volckens, J. AirPen: A Wearable Monitor for Characterizing Exposures to Particulate Matter and Volatile Organic Compounds. *Environ. Sci. Technol.* **2023**, *57* (29), 10604–10614. https://doi.org/10.1021/acs.est.3c02238.

(20) Anderson, J. O.; Thundiyil, J. G.; Stolbach, A. Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health. *J. Med. Toxicol.* **2012**, *8* (2), 166–175. https://doi.org/10.1007/s13181-011-0203-1.

(21) Yang, Y.; Ruan, Z.; Wang, X.; Yang, Y.; Mason, T. G.; Lin, H.; Tian, L. Short-Term and Long-Term Exposures to Fine Particulate Matter Constituents and Health: A Systematic Review and Meta-Analysis. *Environ. Pollut.* **2019**, *247*, 874–882. https://doi.org/10.1016/j.envpol.2018.12.060.

(22) Brown, J. S.; Gordon, T.; Price, O.; Asgharian, B. Thoracic and Respirable Particle Definitions for Human Health Risk Assessment. *Part. Fibre Toxicol.* **2013**, *10* (12), 1–12. https://doi.org/https://doi.org/10.1186/1743-8977-10-12.

(23) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.;

Smith, S. C.; Whitsel, L.; Kaufman, J. D. Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement from the American Heart Association. *Circulation* **2010**, *121* (21), 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1.

(24) Kelly, F. J.; Fussell, J. C. Air Pollution and Public Health: Emerging Hazards and Improved Understanding of Risk. *Environ. Geochem. Health* **2015**, *37* (4), 631–649. https://doi.org/10.1007/s10653-015-9720-1.

(25) Lowther, S. D.; Jones, K. C.; Wang, X.; Whyatt, J. D.; Wild, O.; Booker, D. Particulate Matter Measurement Indoors: A Review of Metrics, Sensors, Needs, and Applications. *Environ. Sci. Technol.* **2019**, *53* (20), 11644–11656. https://doi.org/10.1021/acs.est.9b03425.

(26) Morawska, L.; Thai, P. K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; Gao, J.; Hagler, G. S. W.; Jayaratne, R.; Kumar, P.; Lau, A. K. H.; Louie, P. K. K.; Mazaheri, M.; Ning, Z.; Motta, N.; Mullins, B.; Rahman, M. M.; Ristovski, Z.; Shafiei, M.; Tjondronegoro, D.; Westerdahl, D.; Williams, R. Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environ. Int.* **2018**, *116* (February), 286–299. https://doi.org/10.1016/j.envint.2018.04.018.

(27) Hagan, D.; Kroll, J. Assessing the Accuracy of Low-Cost Optical Particle Sensors Using a Physics-Based Approach. *Atmos. Meas. Tech. Discuss.* **2020**, 1–36. https://doi.org/10.5194/amt-2020-188.

(28) Tryner, J.; Mehaffy, J.; Miller-Lionberg, D.; Volckens, J. Effects of Aerosol Type and

Simulated Aging on Performance of Low-Cost PM Sensors. *J. Aerosol Sci.* **2020**, *150* (June), 105654. https://doi.org/10.1016/j.jaerosci.2020.105654.

(29) Kuula, J.; Mäkelä, T.; Aurela, M.; Teinilä, K.; Varjonen, S.; González, Ó.; Timonen, H. Laboratory Evaluation of Particle-Size Selectivity of Optical Low-Cost Particulate Matter Sensors. *Atmos. Meas. Tech.* **2020**, *13* (5), 2413–2423. https://doi.org/10.5194/amt-13-2413-2020.

(30) Ouimette, J.; Malm, W.; Schichtel, B.; Sheridan, P.; Andrews, E.; Ogren, J.; Arnott, W. P. Evaluating the PurpleAir Monitor as an Aerosol Light Scattering Instrument. *Atmos. Meas. Tech. Discuss.* **2021**, No. June, 1–35. https://doi.org/10.5194/amt-2021-170.

(31) McFarlane, C.; Raheja, G.; Malings, C.; Appoh, E. K. E.; Hughes, A. F.; Westervelt, D. M. Application of Gaussian Mixture Regression for the Correction of Low Cost PM2.5 Monitoring Data in Accra, Ghana. *ACS Earth Sp. Chem.* **2021**, *5* (9), 2268–2279. https://doi.org/10.1021/acsearthspacechem.1c00217.

(32) Hall, E. S.; States, U.; Protection, E. Reference and Equivalent Methods Used to Measure National Ambient Air Quality Standards (NAAQS) Criteria Air Pollutants - Volume I Authors : Joseph H . Gilliam (EPA / ORD). **2016**, *I* (June). https://doi.org/10.13140/RG.2.1.2423.2563.

(33) United States Environmental Protection Agency. *US-EPA Approval GRIMM EDM180*. https://www.grimm-aerosol.com/fileadmin/files/grimm-aerosol/Certificates/US-EPA_Approval_GRIMM_EDM180_180_.pdf (accessed 2022-06-29).

(34) Sira Certification Service. *PRODUCT CONFORMITY CERTIFICATE GRIMM model EDM 180 / EDM 180+ for PM10 and PM2.5*. https://www.csagroupuk.org/wp-

content/uploads/2018/02/MC12019804.pdf (accessed 2022-06-29).

(35) Pawar, H.; Sinha, B. Humidity, Density, and Inlet Aspiration Efficiency Correction Improve Accuracy of a Low-Cost Sensor during Field Calibration at a Suburban Site in the North-Western Indo-Gangetic Plain (NW-IGP). *Aerosol Sci. Technol.* **2020**, *54* (6), 685–703. https://doi.org/10.1080/02786826.2020.1719971.

(36) Giordano, M. R.; Malings, C.; Pandis, S. N.; Presto, A. A.; McNeill, V. F.; Westervelt, D. M.; Beekmann, M.; Subramanian, R. From Low-Cost Sensors to High-Quality Data: A Summary of Challenges and Best Practices for Effectively Calibrating Low-Cost Particulate Matter Mass Sensors. *J. Aerosol Sci.* **2021**, *158* (July), 105833. https://doi.org/10.1016/j.jaerosci.2021.105833.

(37) Barkjohn, K. K.; Gantt, B.; Clements, A. L. Development and Application of a United States-Wide Correction for PM2.5 Data Collected with the PurpleAir Sensor. *Atmos. Meas. Tech.* **2021**, *14* (6), 4617–4637. https://doi.org/10.5194/amt-14-4617-2021.

(38) Malings, C.; Tanzer, R.; Hauryliuk, A.; Saha, P. K.; Robinson, A. L.; Presto, A. A.; Subramanian, R. Fine Particle Mass Monitoring with Low-Cost Sensors: Corrections and Long-Term Performance Evaluation. *Aerosol Sci. Technol.* **2020**, *54* (2), 160–174. https://doi.org/10.1080/02786826.2019.1623863.

(39) Wallace, L.; Bi, J.; Ott, W. R.; Sarnat, J.; Liu, Y. Calibration of Low-Cost PurpleAir Outdoor Monitors Using an Improved Method of Calculating PM2.5. *Atmos. Environ.* **2021**, *256* (March), 118432. https://doi.org/10.1016/j.atmosenv.2021.118432.

(40) European Commission. *Guide to the Demonstration of Equivalence of Ambient Air Monitoring Methods*. https://www.aces.su.se/reflab/wp-

content/uploads/2016/11/Demonstration_of_Equivalence_of_Ambient_Air_Monitoring.pdf (accessed 2022-11-02).

(41) Kosmopoulos, G.; Salamalikis, V.; Pandis, S. N.; Yannopoulos, P.; Bloutsos, A. A.; Kazantzidis, A. Low-Cost Sensors for Measuring Airborne Particulate Matter: Field Evaluation and Calibration at a South-Eastern European Site. *Sci. Total Environ.* **2020**, *748*, 141396. https://doi.org/10.1016/j.scitotenv.2020.141396.

(42) Kelly, K. E.; Whitaker, J.; Petty, A.; Widmer, C.; Dybwad, A.; Sleeth, D.; Martin, R.; Butterfield, A. Ambient and Laboratory Evaluation of a Low-Cost Particulate Matter Sensor. *Environ. Pollut.* **2017**, *221*, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039.

(43) Mei, H.; Han, P.; Wang, Y.; Zeng, N.; Liu, D.; Cai, Q.; Deng, Z.; Wang, Y.; Pan, Y.; Tang, X. Field Evaluation of Low-Cost Particulate Matter Sensors in Beijing. *Sensors (Switzerland)* **2020**, *20* (16), 1–17. https://doi.org/10.3390/s20164381.

(44) Wallace, L.; Bi, J.; Ott, W. R.; Sarnat, J.; Liu, Y. Calibration of Low-Cost PurpleAir Outdoor Monitors Using an Improved Method of Calculating PM2.5. *Atmos. Environ.* **2021**, *256* (May), 118432. https://doi.org/10.1016/j.atmosenv.2021.118432.

(45) Zimmerman, N. Tutorial: Guidelines for Implementing Low-Cost Sensor Networks for Aerosol Monitoring. *J. Aerosol Sci.* **2022**, *159* (July 2021), 105872. https://doi.org/10.1016/j.jaerosci.2021.105872.

(46) Alfano, B.; Barretta, L.; Del Giudice, A.; De Vito, S.; Di Francia, G.; Esposito, E.; Formisano, F.; Massera, E.; Miglietta, M. L.; Polichetti, T. Correction to: A Review of Low-Cost Particulate Matter Sensors from the Developers' Perspectives (Sensors 2020,

20, 6819). *Sensors* **2021**, *21* (9). https://doi.org/10.3390/s21093060.

(47) OpenAQ. *OpenAQ Data Repository*. https://openaq.org/ (accessed 2023-01-10).

(48) Lee, C. H.; Wang, Y. Bin; Yu, H. L. An Efficient Spatiotemporal Data Calibration Approach for the Low-Cost PM2.5 Sensing Network: A Case Study in Taiwan. *Environ. Int.* **2019**, *130* (June), 104838. https://doi.org/10.1016/j.envint.2019.05.032.

(49) Krebs, B.; Burney, J.; Zivin, J. G.; Neidell, M. Using Crowd-Sourced Data to Assess the Temporal and Spatial Relationship between Indoor and Outdoor Particulate Matter. *Environ. Sci. Technol.* **2021**, *55* (9), 6107–6115. https://doi.org/10.1021/acs.est.0c08469.

(50) Singh, A.; Ng'ang'a, D.; Gatari, M. J.; Kidane, A. W.; Alemu, Z. A.; Derrick, N.; Webster, M. J.; Bartington, S. E.; Thomas, G. N.; Avis, W.; Pope, F. D. Air Quality Assessment in Three East African Cities Using Calibrated Low-Cost Sensors with a Focus on Road-Based Hotspots. *Environ. Res. Commun.* **2021**, *3* (7). https://doi.org/10.1088/2515-7620/ac0e0a.

(51) Jaffe, D. A.; Miller, C.; Thompson, K.; Finley, B.; Nelson, M.; Ouimette, J.; Andrews, E. An Evaluation of the U.S. EPA's Correction Equation for Purple Air Sensor Data in Smoke, Dust and Wintertime Urban Pollution Events. *Atmos. Meas. Tech. Discuss.* **2022**, No. October, 1–20. https://doi.org/https://doi.org/10.5194/amt-2022-265.

(52) Delp, W. W.; Singer, B. C. Wildfire Smoke Adjustment Factors for Low-Cost and Professional Pm2.5 Monitors with Optical Sensors. *Sensors (Switzerland)* **2020**, *20* (13), 1–21. https://doi.org/10.3390/s20133683.

(53) Holder, A. L.; Mebust, A. K.; Maghran, L. A.; McGown, M. R.; Stewart, K. E.; Vallano,

D. M.; Elleman, R. A.; Baker, K. R. Field Evaluation of Low-cost Particulate Matter Sensors for Measuring Wildfire Smoke. *Sensors (Switzerland)* **2020**, *20* (17), 1–17. https://doi.org/10.3390/s20174796.

(54) Danek, T.; Zaręba, M. The Use of Public Data from Low-Cost Sensors for the Geospatial Analysis of Air Pollution from Solid Fuel Heating during the COVID-19 Pandemic Spring Period in Krakow, Poland. *Sensors* **2021**, *21* (15). https://doi.org/10.3390/s21155208.

(55) Curto, A.; Donaire-Gonzalez, D.; Barrera-Gómez, J.; Marshall, J. D.; Nieuwenhuijsen, M. J.; Wellenius, G. A.; Tonne, C. Performance of Low-Cost Monitors to Assess Household Air Pollution. *Environ. Res.* **2018**, *163* (October 2017), 53–63. https://doi.org/10.1016/j.envres.2018.01.024.

(56) World Health Organization; International Labour Organization. *WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury, 2000-2016*; Geneva, 2021. https://doi.org/10.1016/j.shaw.2021.12.713.

(57) Pega, F.; Hamzaoui, H.; Náfrádi, B.; Momen, N. C. Global, Regional and National Burden of Disease Attributable to 19 Selected Occupational Risk Factors for 183 Countries, 2000–2016: A Systematic Analysis from the WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury. *Scand. J. Work. Environ. Heal.* **2022**, *48* (2), 158–168. https://doi.org/10.5271/sjweh.4001.

(58) Nelson, D. I.; Concha-Barrientos, M.; Driscoll, T.; Steenland, K.; Fingerhut, M.; Punnett, L.; Prüss-Üstün, A.; Leigh, J.; Corvalan, C. The Global Burden of Selected Occupational Diseases and Injury Risks: Methodology and Summary. *Am. J. Ind. Med.* **2005**, *48* (6), 400–418. https://doi.org/10.1002/ajim.20211.

(59) Teixeira, L. R.; Pega, F.; de Abreu, W.; de Almeida, M. S.; de Andrade, C. A. F.; Azevedo, T. M.; Dzhambov, A. M.; Hu, W.; Macedo, M. R. V.; Martínez-Silveira, M. S.; Sun, X.; Zhang, M.; Zhang, S.; Correa da Silva, D. T. The Prevalence of Occupational Exposure to Noise: A Systematic Review and Meta-Analysis from the WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury. *Environ. Int.* **2021**, *154*, 106380. https://doi.org/10.1016/j.envint.2021.106380.

(60) Gopinath, B.; McMahon, C.; Tang, D.; Burlutsky, G.; Mitchell, P. Workplace Noise Exposure and the Prevalence and 10-Year Incidence of Age-Related Hearing Loss. *PLoS One* **2021**, *16* (7 July), 1–10. https://doi.org/10.1371/journal.pone.0255356.

(61) Gan, W. Q.; Davies, H. W.; Demers, P. A. Exposure to Occupational Noise and Cardiovascular Disease in the United States: The National Health and Nutrition Examination Survey 1999-2004. *Occup. Environ. Med.* **2011**, *68* (3), 183–190. https://doi.org/10.1136/oem.2010.055269.

(62) Yang, Y.; Zhang, E.; Zhang, J.; Chen, S.; Yu, G.; Liu, X.; Peng, C.; Lavin, M. F.; Du, Z.; Shao, H. Relationship between Occupational Noise Exposure and the Risk Factors of Cardiovascular Disease in China A Meta-Analysis. *Med. (United States)* **2018**, *97* (30). https://doi.org/10.1097/MD.0000000000011720.

(63) Skogstad, M.; Johannessen, H. A.; Tynes, T.; Mehlum, I. S.; Nordby, K. C.; Lie, A. Systematic Review of the Cardiovascular Effects of Occupational Noise. *Occup. Med. (Chic. Ill).* **2016**, *66* (1), 10–16. https://doi.org/10.1093/occmed/kqv148.

(64) Themann, C. L.; Masterson, E. A. Occupational Noise Exposure: A Review of Its Effects, Epidemiology, and Impact with Recommendations for Reducing Its Burden. *J. Acoust.*

Soc. Am. **2019**, *146* (5), 3879–3905. https://doi.org/10.1121/1.5134465.

(65) Yoon, J. H.; Won, J. U.; Lee, W.; Jung, P. K.; Roh, J. Occupational Noise Annoyance Linked to Depressive Symptoms and Suicidal Ideation: A Result from Nationwide Survey of Korea. *PLoS One* **2014**, *9* (8). https://doi.org/10.1371/journal.pone.0105321.

(66) Leigh, J. P.; Romano, P. S.; Schenker, M. B.; Kreiss, K. Costs of Occupational COPD and Asthma. *Chest* **2002**, *121* (1), 264–272. https://doi.org/10.1378/chest.121.1.264.

(67) Leigh, J. P. Economic Burden of Occupational Injury and Illness in the United States. *Milbank Q.* **2011**, *89* (4), 728–772. https://doi.org/10.1111/j.1468-0009.2011.00648.x.

(68) Si, S.; Lewkowski, K.; Fritschi, L.; Heyworth, J.; Liew, D.; Li, I. Productivity Burden of Occupational Noise-Induced Hearing Loss in Australia: A Life Table Modelling Study. *Int. J. Environ. Res. Public Health* **2020**, *17* (13), 1–9. https://doi.org/10.3390/ijerph17134667.

(69) Lavoue, J.; Friesen, M. C.; Burstyn, I. Workplace Measurements by the Us Occupational Safety and Health Administration since 1979: Descriptive Analysis and Potential Uses for Exposure Assessment. *Ann. Occup. Hyg.* **2013**, *57* (1), 77–97. https://doi.org/10.1093/annhyg/mes055.

(70) Bureau of Labour Statistics; U.S. Department of Labour. *Occupational Employment and Wage Statistics*. www.bls.gov/oes/ (accessed 2024-09-23).

(71) Hall, A. L.; Peters, C. E.; Demers, P. A.; Davies, H. W. Exposed! Or Not? The Diminishing Record of Workplace Exposure in Canada. *Can. J. Public Heal.* **2014**, *105* (3), 2–5. https://doi.org/10.17269/cjph.105.4266.

(72) Stamm, R. Mega-Database: One Million Data since 1972. *Appl. Occup. Environ. Hyg.* **2001**, *16* (2), 159–163. https://doi.org/10.1080/104732201460262.

(73) UK Health and Safety Executive. *National Exposure Database*. https://www.data.gov.uk/dataset/afc1ad32-f7f6-4f54-a8fc-2a281a25b8ad/national-exposure-database (accessed 2024-09-26).

(74) Mater, G.; Paris, C.; Lavoué, J. Descriptive Analysis and Comparison of Two French Occupational Exposure Databases: COLCHIC and SCOLA. *Am. J. Ind. Med.* **2016**, *59* (5), 379–391. https://doi.org/10.1002/ajim.22569.

(75) Oldham, P. D. The Nature of the Variability of Dust Concentrations at the Coal Face. *Br. J. Ind. Med.* **1953**, *10* (4), 227.

(76) Nothstein, G. L.; Hahne, R. M. A.; Spence, M. W. Evaluation of the Cost-Effectiveness of Various Passive Monitors and Active Monitors for Industrial Hygiene Sampling. *AIHAJ - Am. Ind. Hyg. Assoc.* **2000**, *61* (1), 64–68. https://doi.org/10.1080/15298660008984516.

(77) Schall, M. C.; Sesek, R. F.; Cavuoto, L. A. Barriers to the Adoption of Wearable Sensors in the Workplace: A Survey of Occupational Safety and Health Professionals. *Hum. Factors* **2018**, *60* (3), 351–362. https://doi.org/10.1177/0018720817753907.

(78) NIOSH. *Particulates Not Otherwise Regulated, Total: Method 0500*; 1994; pp 1–3. https://www.cdc.gov/niosh/docs/2003-154/pdfs/0500.pdf.

(79) L'Orange, C.; Anderson, K.; Sleeth, D.; Anthony, T. R.; Volckens, J. A Simple and Disposable Sampler for Inhalable Aerosol. *Ann. Occup. Hyg.* **2015**, *60* (2), 150–160. https://doi.org/10.1093/annhyg/mev065.

(80) Kenny, L. C.; Aitken, R. J.; Baldwin, P. E. J.; Beaumont, G. C.; Maynard, A. D. The Sampling Efficiency of Personal Inhalable Aerosol Samplers in Low Air Movement Environments. *J. Aerosol Sci.* **1999**, *30* (5), 627–638. https://doi.org/10.1016/S0021-8502(98)00752-6.

(81) Görner, P.; Simon, X.; Wrobel, R.; Kauffer, E.; Witschger, O. Laboratory Study of Selected Personal Inhalable Aerosol Samplers. *Ann. Occup. Hyg.* **2010**, *54* (2), 165–187. https://doi.org/10.1093/annhyg/mep079.

(82) US Department of Transportation. Lithium Battery Guide for Shippers. **2021**, No. September, 1–36.

(83) L'orange, C.; Neymark, G.; Carter, E.; Volckens, J. A High-Throughput, Robotic System for Analysis of Aerosol Sampling Filters. *Aerosol Air Qual. Res.* **2021**, *21* (11), 1–14. https://doi.org/10.4209/AAQR.210037.

(84) Fedak, K. M.; Good, N.; Dahlke, J.; Hecobian, A.; Sullivan, A.; Zhou, Y.; Peel, J. L.; Volckens, J. Chemical Composition and Emissions Factors for Cookstove Startup (Ignition) Materials. *Environ. Sci. Technol.* **2018**. https://doi.org/10.1021/acs.est.8b02218.

(85) Bilsback, K. R.; Dahlke, J.; Fedak, K. M.; Good, N.; Hecobian, A.; Herckes, P.; Orange, C. L.; Meha, J.; Sullivan, A.; Tryner, J.; Zyl, L. Van; Walker, E. S.; Zhou, Y.; Pierce, R.; Wilson, A.; Peel, J. L.; Volckens, J. A Laboratory Assessment of 120 Air Pollutant Emissions from Biomass and Fossil Fuel Cookstoves. *Environ. Sci. Technol.* **2019**. https://doi.org/10.1021/acs.est.8b07019.

(86) Zyl, L. Van; Tryner, J.; Bilsback, K. R.; Good, N.; Hecobian, A.; Sullivan, A.; Zhou, Y.; Peel, J. L.; Volckens, J. Effects of Fuel Moisture Content on Emissions from a Rocket-

Elbow Cookstove. *Environ. Sci. Technol.* **2019**. https://doi.org/10.1021/acs.est.9b00235.

(87) International Electrotechnical Commission. IEC 61672-1 International Standard: Electroacoustics - Sound Level Meters - Part 1: Specifications. Geneva 2013.

(88) Johnson, D. L.; Bell, M. L. Sources and Control of Error in Industrial Hygiene Measurements. **1991**, *39* (8), 362–368.

(89) TIELEMANS, E.; MARQUART, H.; DE COCK, J.; GROENEWOLD, M.; VAN HEMMEN, J. A Proposal for Evaluation of Exposure Data. *Ann. Occup. Hyg.* **2002**, *46* (3), 287–297. https://doi.org/10.1093/annhyg/mef047.

(90) Lawless, P.; Thornburg, J.; Rodes, C.; Williams, R. Personal Exposure Monitoring Wearing Protocol Compliance: An Initial Assessment of Quantitative Measurement. *J. Expo. Sci. Environ. Epidemiol.* **2012**, *22* (3), 274–280. https://doi.org/10.1038/jes.2012.8.

(91) Pleil, J. D. Imputing Defensible Values for Left-Censored ‘ below Level of Quantitation ’ (LoQ) Biomarker Measurements. *J. Breath Res.* **2023**, No. October 2016, 45001. https://doi.org/10.1088/1752-7155/10/4/045001.

(92) Venables, W. N.; Ripley, B. D. *Modern Applied Statistics with S*, Fourth.; Springer: New York, 2002.

(93) Cullen, A. C.; Frey, H. C. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, 1st ed.; Language of science; Springer US: New York, 1999.

(94) Natrella, M. G. *Experimental Statistics*; Experimental Statistics; U.S. Department of Commerce, National Bureau of Standards, 1963.

(95) Ramachandran, G. *Occupational Exposure Assessment for Air Contaminants*, 1st ed.; CRC Press: Boca Raton, 2005.

(96) Occupational Safety and Health Administration. *Occupational Safety and Health Standards - Toxic and Hazardous Substances - Air Contaminants*; United States.

(97) Barsan, M. E. NIOSH Pocket Guide to Chemical Hazards. *DHHS publication ; (NIOSH)*. Cincinnati, OH 2005. https://stacks.cdc.gov/view/cdc/21265.

(98) Occupational Safety and Health Administration. *Occupational Safety and Health Standards - Occupational Noise Exposure*; United States.

(99) Gardiner, K. Needs of Occupational Exposure Sampling Strategies for Compliance and Epidemiology. *Occup. Environ. Med.* **1995**, *52* (11), 705–708. https://doi.org/10.1136/oem.52.11.705.

(100) Stokholm, Z. A.; Hansen, Å. M.; Grynderup, M. B.; Bonde, J. P.; Christensen, K. L.; Frederiksen, T. W.; Lund, S. P.; Vestergaard, J. M.; Kolstad, H. A. Recent and Long-Term Occupational Noise Exposure and Salivary Cortisol Level. *Psychoneuroendocrinology* **2014**, *39* (1), 21–32. https://doi.org/10.1016/j.psyneuen.2013.09.028.

(101) Sayler, S. K.; Long, R. N.; Nambunmee, K.; Neitzel, R. L. Respirable Silica and Noise Exposures among Stone Processing Workers in Northern Thailand. *J. Occup. Environ. Hyg.* **2018**, *15* (2), 117–124. https://doi.org/10.1080/15459624.2017.1393080.

(102) Lake, K.; Zhu, J.; Wang, H.; Volckens, J.; Koehler, K. A. Effects of Data Sparsity and Spatiotemporal Variability on Hazard Maps of Workplace Noise. *J. Occup. Environ. Hyg.* **2015**, *12* (4), 256–265. https://doi.org/10.1080/15459624.2014.963589.

(103) Hewett, P. Sample Size Formulae for Estimating the True Arithmetic or Geometric Mean of Lognormal Exposure Distributions. *Am. Ind. Hyg. Assoc. J.* **1995**, *56* (3), 219–225. https://doi.org/10.1080/15428119591017042.

(104) SELVIN, S.; RAPPAPORT, S.; SPEAR, R.; SCHULMAN, J.; FRANCIS, M. A Note on the Assessment of Exposure Using One-Sided Tolerance Limits. *Am. Ind. Hyg. Assoc. J.* **1987**, *48* (2), 89–93. https://doi.org/10.1080/15298668791384445.

(105) Leidel, N. A.; Busch, K. A.; Lynch, J. R.; for Occupational Safety, N. I.; Health; Systems Control, inc. *Occupational Exposure Sampling Strategy Manual*; DHEW publication; Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health, 1977.

(106) Suh, H. H.; Bahadori, T.; Vallarino, J.; Spengler, J. D. Criteria Air Pollutants and Toxic Air-Pollutants. *Environ. Health Perspect.* **2000**, *108* (SUPPL. 4), 625–633. https://doi.org/10.1289/ehp.00108s4625.

(107) United States Environmental Protection Agency. *National Air Toxics Program : The Second Integrated Urban Air Toxics Report to Congress*; Research Triangle Park, 2014. http://www2.epa.gov/sites/production/files/2014-08/documents/082114-urban-air-toxics-report-congress.pdf.

(108) Bradstreet, J. W. *Hazardous Air Pollutants: Assessment, Liabilities and Regulatory Compliance*; William Andrew, 1996.

(109) United States Environmental Protection Agency. *2020 National Emissions Inventory and Trends Report*; Research Triangle Park, 2023. https://storymaps.arcgis.com/stories/d7d730f974c6474190b142a49ae8d3bd.

(110) Watkins, N. *Ambient Air Toxics Monitoring Program Overview*; 2022.

(111) United States Environmental Protection Agency. *Appendix D to Part 58*; Code of Federal Regulations, 2024; Vol. 40, pp 1–25. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-58/appendix-Appendix D to Part 58.

(112) United States Environmental Protection Agency. Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air, 2nd. Edition. **1999**, *EPA/625/R*-.

(113) Woolfenden, E. A.; Subramanian, G. Practical Aspects of Monitoring Volatile Organics in Air. In *Quality Assurance in Environmental Monitoring*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 1995; pp 133–182.

(114) Woolfenden, E. Sorbent-Based Sampling Methods for Volatile and Semi-Volatile Organic Compounds in Air: Part 1: Sorbent-Based Air Monitoring Options. *J. Chromatogr. A* **2010**, *1217* (16), 2674–2684. https://doi.org/https://doi.org/10.1016/j.chroma.2009.12.042.

(115) Strum, M.; Scheffe, R. National Review of Ambient Air Toxics Observations. *J. Air Waste Manag. Assoc.* **2016**, *66* (2), 120–133. https://doi.org/10.1080/10962247.2015.1076538.

(116) Hersher, R. Millions Of Pounds Of Extra Pollution Were Released Before Hurricane Laura's Landfall. *National Public Radio*. August 28, 2020. https://www.npr.org/sections/health-shots/2020/08/28/906822940/millions-of-pounds-of-extra-pollution-were-released-before-laura-made-landfall.

(117) Chism, C. *Air monitoring during hurricanes*. We Were Wondering: A TCEQ Blog.

https://blog.tceq.texas.gov/2021/06/02/air-monitoring-during-hurricanes/.

(118) United States Environmental Protection Agency. *EPA Needs to Improve Its Emergency Planning to Better Address Air Quality Concerns During Future Disasters OFFICE OF INSPECTOR GENERAL Improving Air Quality*; Washington D.C., 2019. www.epa.gov/oig.

(119) Formaldehyde, 2-Butoxyethanol and 1-Tert-Butoxypropan-2-Ol. *IARC Monogr. Eval. Carcinog. risks to humans* **2006**, *88*, 1–478.

(120) Houston Health Department. *Fifth Ward/Kashmere Gardens Union Pacific Railroad Site Contamination And Area Cancer Cluster*. https://www.houstonhealth.org/services/pollution-control/fifth-ward-kashmere-gardens-union-pacific-railroad-site-contamination-area-cancer-cluster (accessed 2025-06-11).

(121) Sansom, G. T.; Fawkes, L. S.; Thompson, C. M.; Losa, L. M.; McDonald, T. J.; Chiu, W. A. Cancer Risk Associated with Soil Distribution of Polycyclic Aromatic Hydrocarbons within Three Environmental Justice Neighborhoods in Houston, Texas. *Environ. Geochem. Health* **2023**, *45* (2), 333–342. https://doi.org/10.1007/s10653-022-01245-5.

(122) Phillips, A. *Formaldehyde Air Pollution in Houston*; Houston, 2021. https://environmentalintegrity.org/wp-content/uploads/2021/07/Houston-Formaldehyde-Report-Final-2021.pdf.

(123) Uchiyama, S.; Otsubo, Y. Simultaneous Determination of Ozone and Carbonyls Using Trans-1,2-Bis(4-Pyridyl)Ethylene as an Ozone Scrubber for 2,4- Dinitrophenylhydrazine-Impregnated Silica Cartridge. *Anal. Chem.* **2008**, *80* (9), 3285–3290. https://doi.org/10.1021/ac702475s.

(124) Texas Commission on Environmental Quality. *Texas Air Monitoring Information System (TAMIS) Web Interface*. https://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=home.welcome (accessed 2025-06-11).

(125) Texas Commission on Environmental Quality. *2024 Annual Monitoring Network Plan*; 2024.

(126) Haney, J. T. *Development Support Document - Formaldehyde, CAS Registry Number: 50-00-0*; Austin, 2015. https://www.tceq.texas.gov/downloads/toxicology/dsd/final/formaldehyde.pdf.

(127) United States Environmental Protection Agency. *IRIS Toxicological Review of Formaldehyde (Inhalation)*; Washington D.C., 2024. http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=497038.

(128) Salthammer, T. Formaldehyde in the Ambient Atmosphere: From an Indoor Pollutant to an Outdoor Pollutant? *Angew. Chemie Int. Ed.* **2013**, *52* (12), 3320–3327. https://doi.org/https://doi.org/10.1002/anie.201205984.

(129) United States Environmental Protection Agency. *Method TO-11A: Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air Second Edition Compendium Method TO-11A Determination of Formaldehyde in Ambient Air Using Adsorbent Cartridge Followed by High Performance Liquid Chromat*; 1999.

(130) Agency for Toxic Substances and Disease Registry. Public Health Assessment Guidance Manual. Public Health Service: Atlanta 2005, pp 1–357.

(131) Texas Deparment of State Health Services. *Assessment of the Occurrence of Cancer Supplemental Assessment*; Houston, 2021.

(132) Texas Deparment of State Health Services. *Assessment of the Occurrence of Cancer in Houston, Texas (2000 - 2016)*; Houston, 2020.

(133) Li, Y.; Zhao, Y.; Kleeman, M. J. Formaldehyde Exposure Racial Disparities in Southeast Texas. *Environ. Sci. Technol.* **2024**, *58* (10), 4680–4690. https://doi.org/10.1021/acs.est.3c02282.

(134) Yong, Z.; Plantower. *Plantower PMS5003 Data Manual*. http://www.aqmd.gov/docs/default-source/aq-spec/resources-page/plantower-pms5003-manual_v2-3.pdf (accessed 2022-09-07).

(135) Sensirion. *Datasheet SPS30*. https://sensirion.com/media/documents/8600FF88/616542B5/Sensirion_PM_Sensors_Datasheet_SPS30.pdf (accessed 2022-09-07).

(136) Piera Systems. *Photon Counting Intelligent Particle Sensor for Accurate Air Quality Monitoring Product Specification Product Summary*. https://pierasystems.com/wp-content/uploads/2021/12/IPS-Datasheet-V1.1.8.pdf (accessed 2022-06-29).

(137) Men, Y.; Li, J.; Liu, X.; Li, Y.; Jiang, K.; Luo, Z.; Xiong, R.; Cheng, H.; Tao, S.; Shen, G. Contributions of Internal Emissions to Peaks and Incremental Indoor PM2.5 in Rural Coal Use Households. *Environ. Pollut.* **2021**, *288* (June), 117753. https://doi.org/10.1016/j.envpol.2021.117753.

(138) Li, J.; Men, Y.; Liu, X.; Luo, Z.; Li, Y.; Shen, H.; Chen, Y.; Cheng, H.; Shen, G.; Tao, S.

Field-Based Evidence of Changes in Household PM2.5 and Exposure during the 2020 National Quarantine in China. *Environ. Res. Lett.* **2021**, *16* (9). https://doi.org/10.1088/1748-9326/ac1014.

(139) Cakmak, S.; Dales, R.; Leech, J.; Liu, L. The Influence of Air Pollution on Cardiovascular and Pulmonary Function and Exercise Capacity: Canadian Health Measures Survey (CHMS). *Environ. Res.* **2011**, *111* (8), 1309–1312. https://doi.org/10.1016/j.envres.2011.09.016.

(140) Atallah, L.; Lo, B.; King, R.; Yang, G. Sensor Placement for Activity Detection Using Wearable Accelerometers. *2010 Int. Conf. Body Sens. Networks* **2010**, 24–29. https://doi.org/10.1109/BSN.2010.23.

(141) Jalal, A. A Triaxial Acceleration-Based Human Motion Detection for Ambient Smart Home System. *2019 16th Int. Bhurban Conf. Appl. Sci. Technol.* **2019**, 353–358.

(142) Yee, T. W. *Vector Generalized Linear and Additive Models: With an Implementation in R*; Springer: New York, USA, 2015.

APPENDIX A

1. Low-cost sensor descriptions and specifications

All three sensors utilize light-scattering to estimate aerosol concentrations, including a focused laser diode to illuminate particles that pass through a small sensing zone and a photodiode (placed near the sensing zone but orthogonal to the direction of laser light propagation) to detect the intensity scattered light across a truncated solid angle. Air flow is drawn through each device by a miniature induction fan.

The Plantower PMS5003 is described by the manufacturer as a "digital universal particle concentration sensor".[134] This sensor has been used in multiple commercial low-cost air quality monitors, including the popular PurpleAir (older versions of the PurpleAir used the PMS1003 and newer versions use the PMS6003, but most units including the ones we evaluated use the PMS5003). The Sensirion SPS30 is described by the manufacturer as a particulate matter sensor for air quality monitoring.[135] Piera describes the IPS-7100 as a "photon counting intelligent particle sensor", and their technology as "intuitive direct-particle counting".[136] The IPS-7100 firmware is user-upgradable, and we updated from v1.9.9 to v1.9.10 on Piera's request before the start of our first (fall/winter) data collection period.

We chose these sensors for three reasons: (1) the technology featured in these devices, while similar across the three models, represents the state-of-the-art for commercial low-cost (<$100 USD) aerosol sensing; (2) the selected sensors are from major suppliers that span three continents: Asia (Plantower), North America (Piera), and Europe (Sensirion); and (3) these sensors (especially the Plantower PMS5003) represent the majority of in-use technologies across the world. For example, the Plantower has been deployed widely and in large numbers in

networks throughout countries in Asia, like China and Taiwan,[137–139] where it has a large share of the PM sensor market. The SPS30 and IPS-7100 were selected because they provide geographic diversity (they are from companies in North America and Europe, respectively).

2. Datalogging

We used a PurpleAir PA-II to collect data from two PMS5003 sensors. Logs in CSV format were downloaded from PurpleAir's website. We developed a custom data logger and housing to collect data from two SPS30 sensors (Figure A1). The data logger used a Particle Boron LTE microcontroller to read the sensors' data and send it via a cellular network to a server hosted by the Center for Energy Development and Health (CEDH) at Colorado State University. We collected data from two IPS-7100 sensors and sent those data to the server using a Wi-Fi enabled Espressif ESP32 microcontroller. We made a custom housing for the Piera IPS-7100 sensors similar to what we developed for the SPS30 sensors.

For the Plantower PMS5003 sensor, the "CF=1, standard particle" output variables were used in all calculations involving $PM_{1.0}$, $PM_{2.5}$, and/or PM_{10} concentrations. The "CF=1" variables from the PMS5003 are uncorrected, according to the manufacturer, as opposed to the "ATM" variables which use a proprietary algorithm and calibration unspecified by the manufacturer to estimate atmospheric aerosol concentrations.

One-minute averages of all reported outputs were logged for the SPS30 and IPS-7100 sensors, whereas the PMS5003 (which was operated in a PurpleAir device) logged 2-minute averages. Unless otherwise stated, raw sensor mass and number readings were recorded directly, and only time-based averaging was applied.

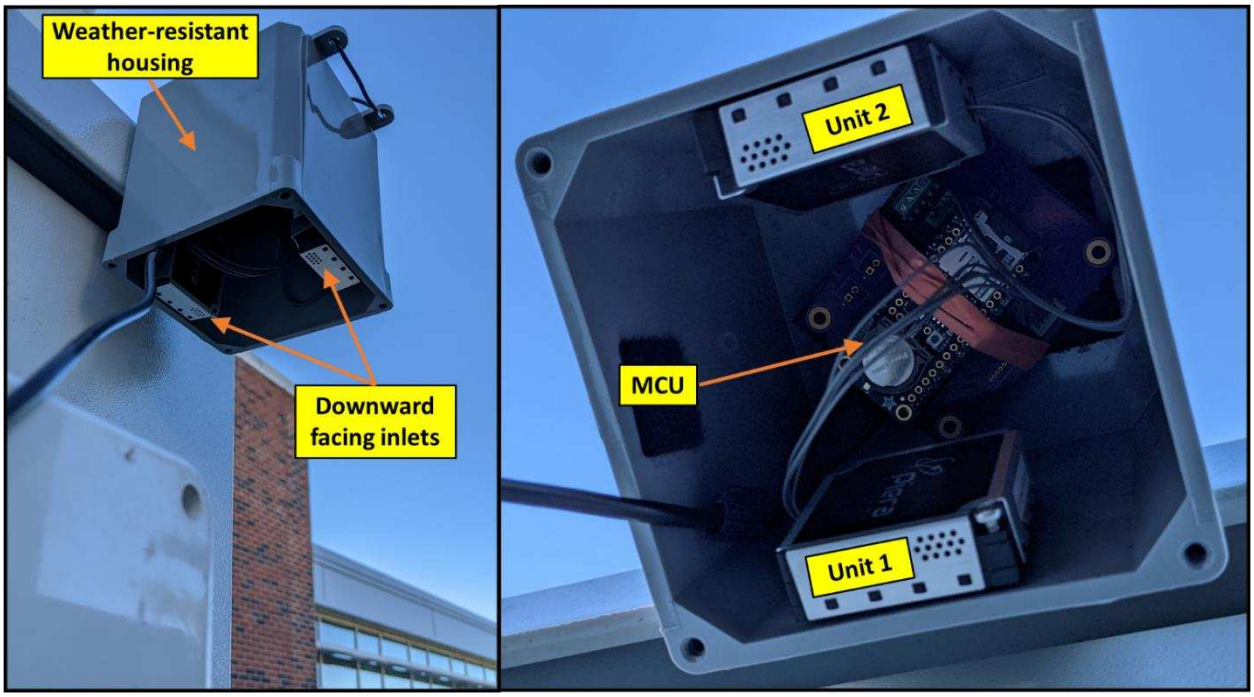

Figure A1: Data logger and housing used for the IPS-7100 (shown in picture) and SPS30 sensors.

3. Common variables

PM_x : Cumulative particulate matter concentration up to the particle size indicated by the subscript (in µm).

$PM_{a \to b}$: Differential particulate matter concentration within the size range indicated by the subscript (in µm).

$PM_{a \to b}$: Also defined as **$PM_b - PM_a$** .

4. Statistical analyses

Descriptive statistics were calculated for each sensor. We present the statistical metrics corresponding to a single unit (i.e., "unit a") of each sensor model, except for the metrics that measure intra-model (unit-to-unit) variability. Standard statistical metrics were calculated to assess sensor precision (coefficient of variation among co-located devices of the same model),

linearity (coefficient of determination vs. reference), and bias (e.g., RMSE, MAE, NMB vs. reference) as a function of particle size range. Where appropriate, we developed linear regression models between sensor and reference data using ordinary least squares to estimate slope, intercept, and as inputs to estimate relative expanded uncertainty (REU), which has been adopted by the European Commission air quality directive as a measure of low-cost sensor performance relative to reference monitors.[40] Visual diagnostics were used to assess model assumptions (i.e., linearity, normality, and homoscedasticity) (Figures A7 and A8). Time-series and scatter plots were developed to visualize sensor performance as a function of particle size range.

5. Performance metrics (equations)

Root mean square error (RMSE):

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - x_i)^2}{n}}$$

Where n is the number of data pairs, y_i is the i[th] low-cost sensor measurement, and x_i is the i[th] reference monitor measurement.

Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - x_i|$$

Where n is the number of data pairs, y_i is the i[th] low-cost sensor measurement, and x_i is the i[th] reference monitor measurement.

Mean bias error (MBE):

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$$

Where n is the number of data pairs, y_i is the i[th] low-cost sensor measurement, and x_i is the i[th] reference monitor measurement.

Normalized mean bias (NMB):

$$NMB = \frac{\sum_{i=1}^{n}(y_i - x_i)}{\sum_{i=1}^{n} x_i}$$

Where n is the number of data pairs, y_i is the i[th] low-cost sensor measurement, and x_i is the i[th] reference monitor measurement.

Coefficient of variation (CV):

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{\frac{\sum(x_j - \mu)^2}{N}}}{\mu}$$

Where σ is the standard deviation of the measurements of all units, μ is the mean of the measurements of all units, N is the number of sensor units, and x_j is the measurement of the j[th] sensor.

Relative expanded uncertainty (REU):

The European Commission guide to the demonstration of equivalence of ambient air monitoring methods recommends the following equation to estimate the REU of low-cost sensors[40]:

$$REU(y_i) = \frac{2\sqrt{\frac{RSS}{n-2} - u^2(bs, RM) + u^2(bs, s) + [b_0 + (b_1 - 1)x_i]^2}}{y_i}$$

Where n is the number of pairs of collocated data (i.e., reference monitor and LCS), y_i is the low-cost sensor measurement, and x_i is the reference monitor measurement. The slope and intercept of a linear regression of y_i as a function of x_i are given by b_1 and b_0 respectively. RSS is the sum of the squared residuals and is calculated as:

$$RSS = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2$$

$u(bs, RM)$ is the between reference method standard uncertainty. Values of this uncertainty are tabulated for many reference monitors. If not available, it is recommended to use a value of 0.67 μg/m^3. This was the case for our reference monitor.

$u(bs, s)$ is the between LCS standard uncertainty. When multiple units of the same sensor model are tested, $u(bs, s)$ can be estimated as:

$$u(bs, s) = \sqrt{\left(\frac{\sum_{l=1}^{N} \sum_{j=1}^{p} (y_{l,j} - \bar{y}_l)^2}{N(p-1)}\right)}$$

Where $y_{l,j}$ is the measurement of sensor unit j for period l. $\bar{y}_l$ is the mean for period l of all sensor units. N is the number of measurements over time. p is the number of collocated sensor units.

6. Relative expanded uncertainty (REU) plots

The REU plot aids in the qualitative and quantitative analysis of errors. The position of the point on the y-axis is a measure of the bias. The scattering of the points along the y-axis is a measure of noise.

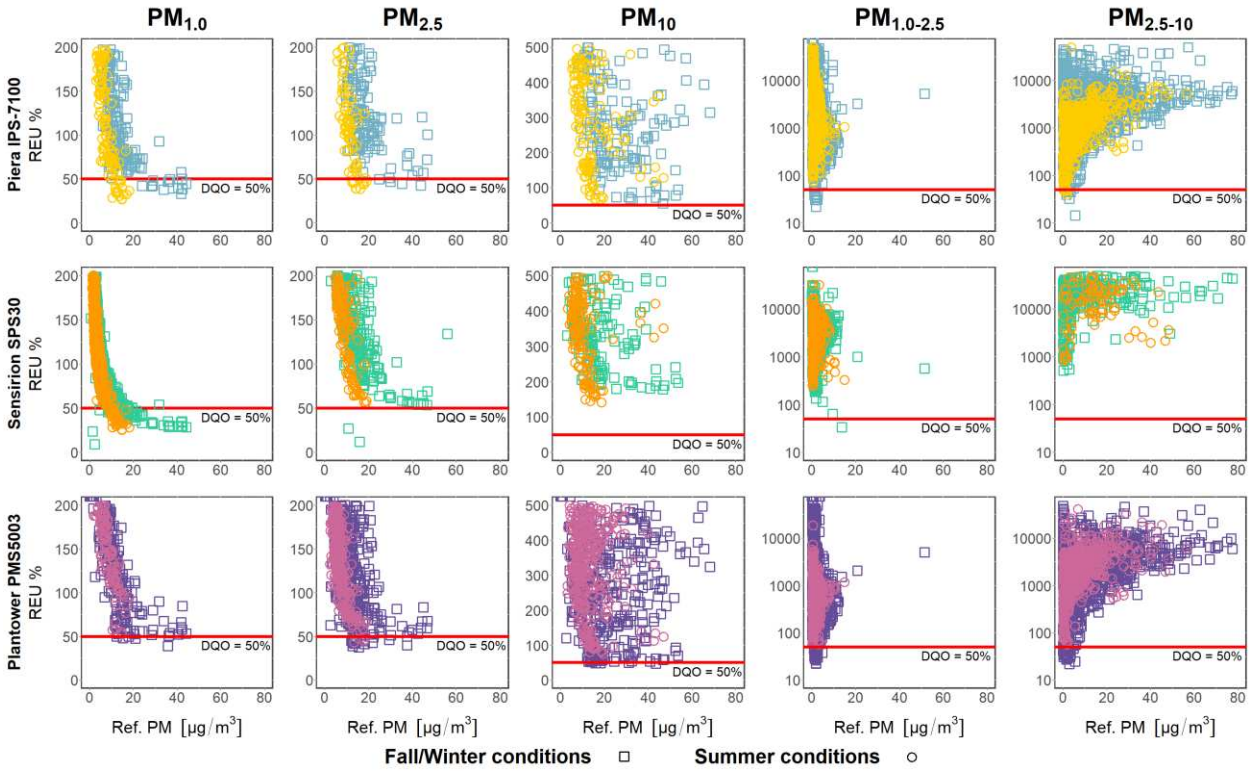

Figure A2: Relative expanded uncertainty (REU) plots for PM estimates from the low-cost sensors. Horizontal axes correspond to the GRIMM EDM180 measurements. The red line is the data quality objective established by the European Commission for low-cost PM sensors (points under the red line have adequate uncertainty). Some points are not shown in the plots due to the axes range.

7. Weather data during the evaluation

The first testing period (23-Nov-2021 to 09-Jan-2022) included late fall and early winter conditions with 13 inches of snowfall. Temperature ranged from -22°C to 23°C and relative humidity ranged from 0% to 77% (Figure A3). The second testing period (13-Jun-2021 to 30-Jul-2022) was comprised of summer conditions with temperature ranging from 11°C to 38°C and relative humidity ranging from 0% to 82% (Figure A4).

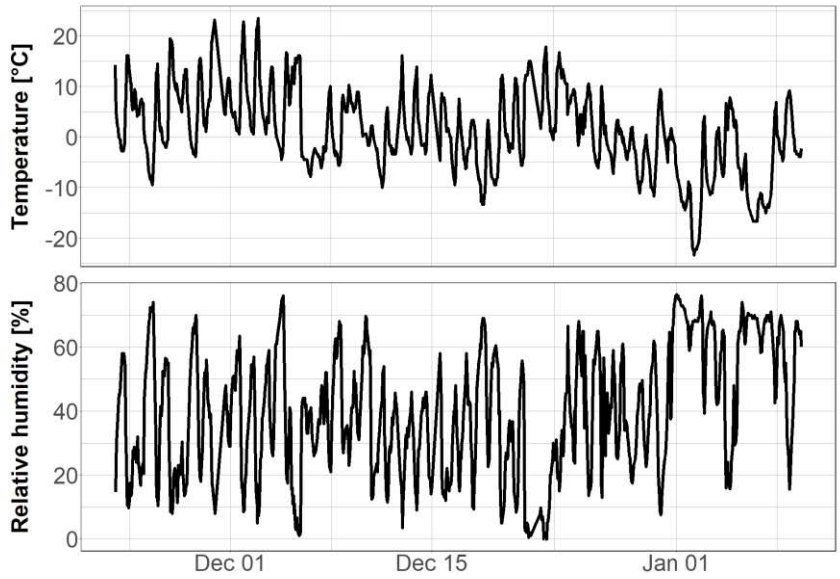

Figure A3: Ambient temperature and relative humidity at 1-hour resolution during the fall/winter campaign.

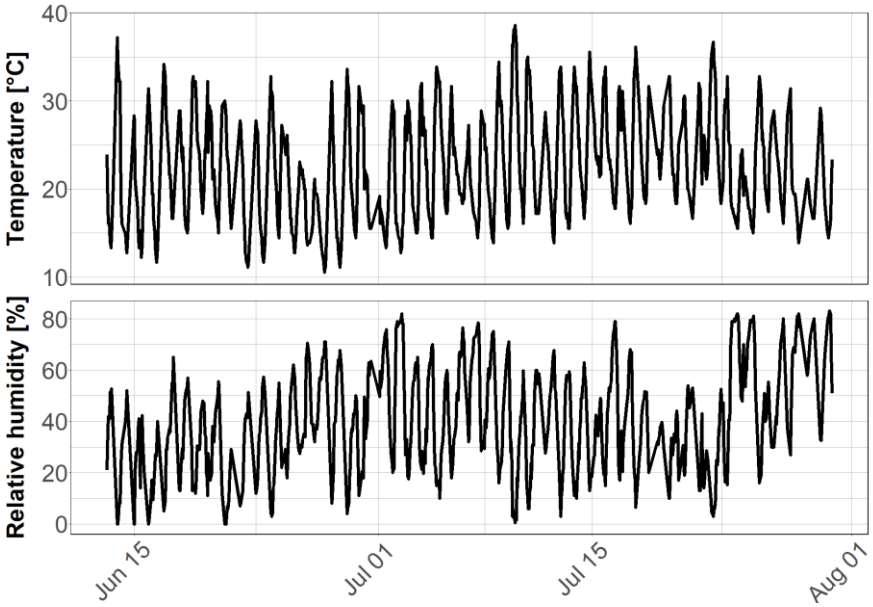

Figure A4: Ambient temperature and relative humidity at 1-hour resolution during the summer campaign.

8. GRIMM EDM180 vs. Thermo Scientific 5014i BAM $PM_{2.5}$ and PM_{10} data for quality assurance

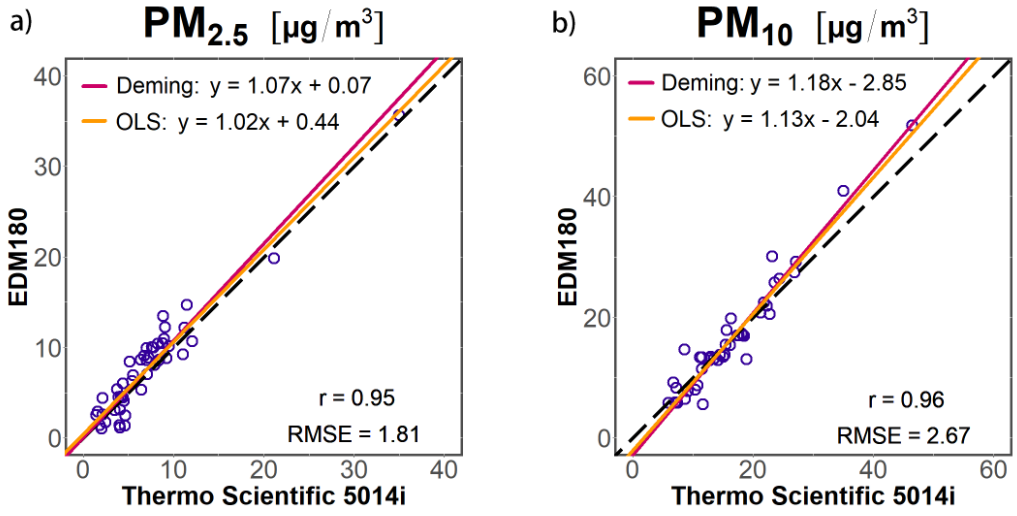

Figure A5: Scatter plots of $PM_{2.5}$ (left) and PM_{10} (right) daily concentrations for the GRIMM EDM180 (reference monitor) and the Thermo Scientific 5014i Beta-attenuation monitor (an FEM monitor). The dashed black line corresponds to the "1:1" relationship. Constant error was assumed for the Deming regression.

9. Time-series plot of the summer period

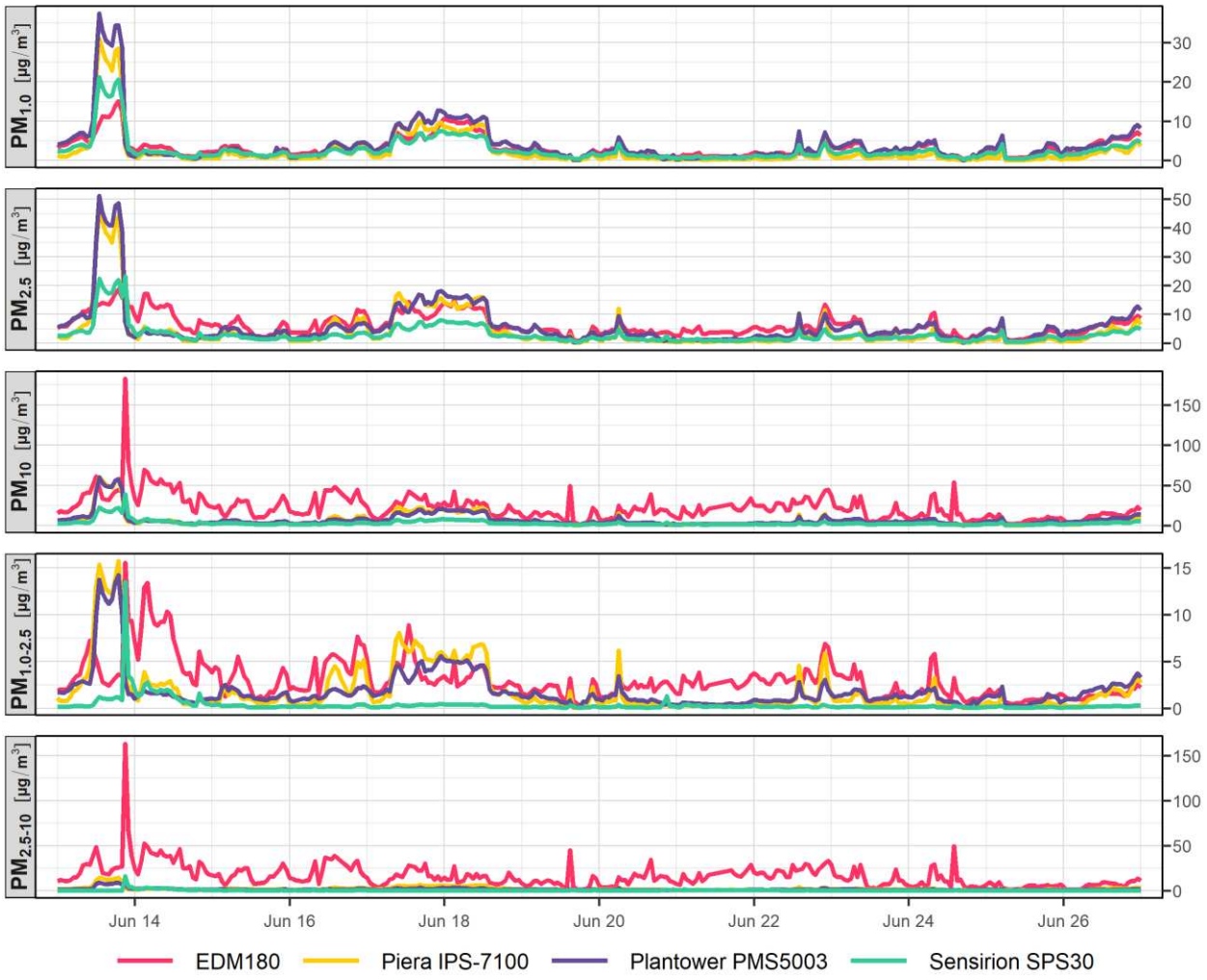

Figure A6: Time series graph of PM concentrations (cumulative and differential) during the first 14 days of the summer period.

10. Confidence intervals of regression model estimates

Table A1. Confidence intervals (2.5% to 97.5%) of the model parameters presented on Table 2. The models are linear regressions of the low-cost sensor vs. the reference monitor (EDM180).

	$PM_{1.0}$	$PM_{2.5}$	PM_{10}	$PM_{1.0-2.5}$	$PM_{2.5-10}$
Slope					
Piera IPS-7100	0.85 to 0.89	1.64 to 1.76	0.17 to 0.25	0.11 to 0.46	0 to 0.015
Sensirion SPS30	0.86 to 0.89	0.75 to 0.80	0.15 to 0.18	0.19 to 0.26	0.05 to 0.06
Plantower PMS5003	1.41 to 1.46	1.62 to 1.73	0.16 to 0.23	0.07 to 0.23	0 to 0.012
Intercept					
Piera IPS-7100	-1.67 to -1.41	-7.04 to -5.98	1.44 to 3.33	1.04 to 2.09	1.45 to 1.73

Sensirion SPS30	-0.62 to -0.45	-1.60 to -1.18	0.46 to 1.22	-0.26 to -0.05	-0.60 to -0.40
Plantower PMS5003	-1.17 to -0.84	-3.41 to -2.61	5.04 to 6.82	1.95 to 2.44	1.36 to 1.68

11. Residual plots of the linear models used in the study

Residual plots are mainly used to evaluate the assumptions of a linear model, such as homoscedasticity, independence, and normality. Qualitatively, a "good model" will have residuals that are uncorrelated (independent), normally distributed and centered around the zero line, and constant with respect to the magnitude of the predicted value (homoscedastic). Residuals that do not follow these patterns suggest one or more violations of the assumptions for a linear model (which is often the case for low-cost sensor data and increasingly evident in the larger size fractions).

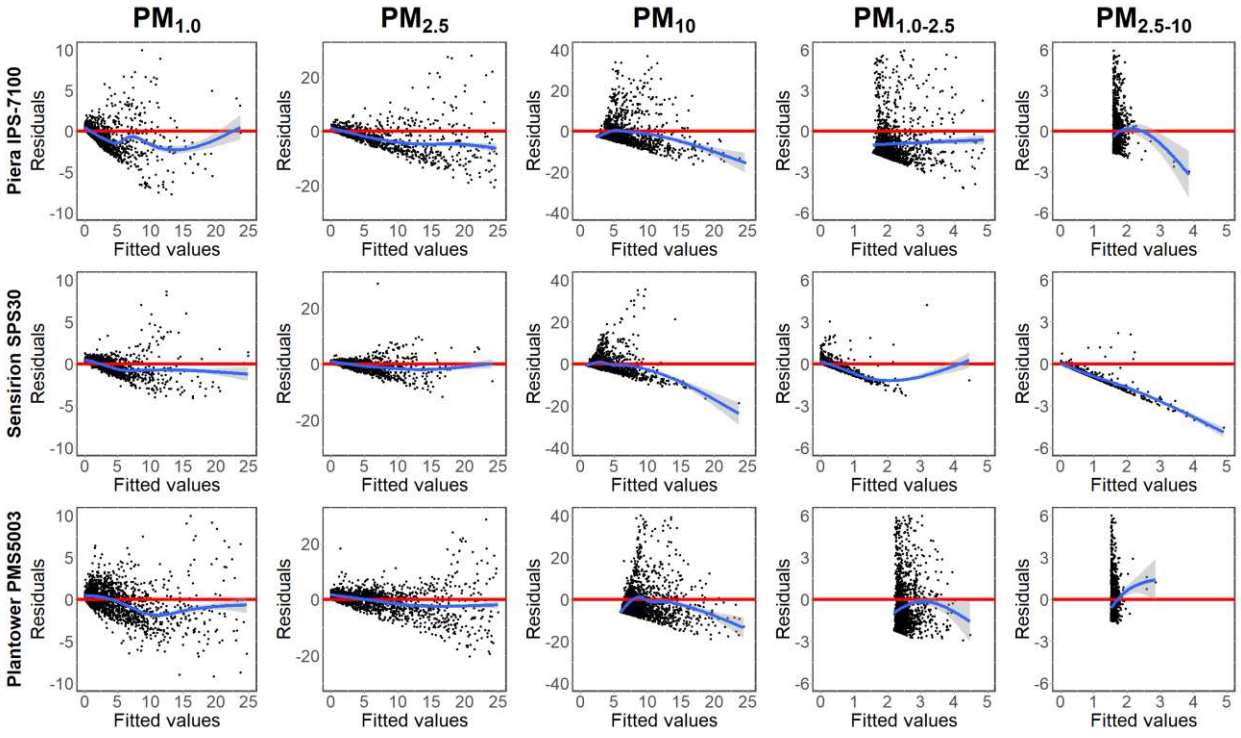

Figure A7: Residual plots corresponding to the low-cost sensor vs. reference monitor linear regression models presented in Figure 4 and mentioned throughout the manuscript. The horizontal red line represents perfect modeled-measured agreement across the range of fitted values; the blue lines represent a LOESS (locally estimated scatterplot smoothing) fit to the observed residuals.

12. Quantile-Quantile plots of the linear models used in the study

Quantile-Quantile plots are used to assess the normality assumption for a linear regression model. The distribution of the residuals (i.e., black dots) is compared against the expected distribution of residuals for an ideal model (i.e., red line). On a "good" Quantile-Quantile plot, the points sit close to the 1:1 line across the data range.

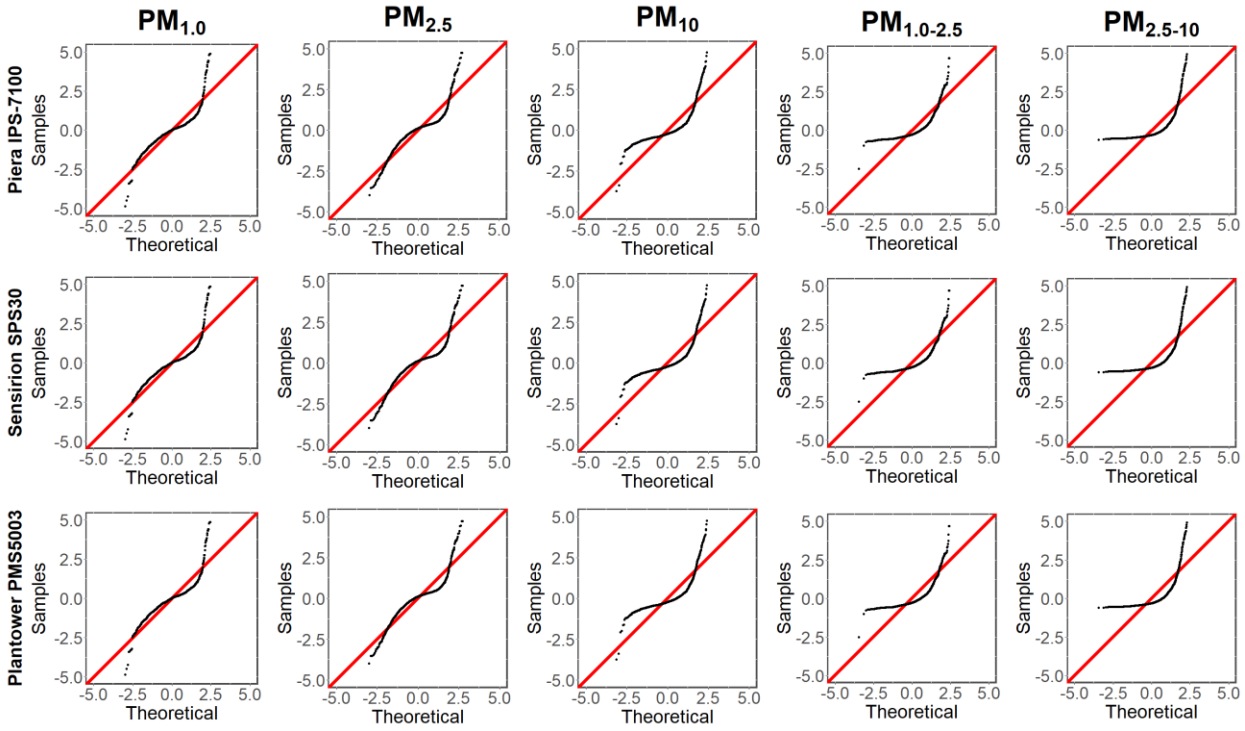

Figure A8: Quantile-Quantile plots corresponding to the low-cost sensor vs. reference monitor linear regression models presented in Figure 4 and mentioned throughout the manuscript.

APPENDIX B

1. Exploratory study results

In December 2022 (approximately one year before the factory-wide sampling campaign), members of the study team paid an initial 2-day visit to the factory to gain a better understanding of the layout and operational schedule as well as the PM, noise, and formaldehyde levels present. Information gleaned during this visit was used to plan the facility-wide campaign. Protocols were designed to collect preliminary data through an exploratory strategy. Area concentrations, as opposed to personal exposure, were measured using stationary samplers and monitors. Prototype AirPens, conventional PM samplers, noise dosimeters, and conventional gas samplers were colocated at several locations across the factory (example in Figure B1).

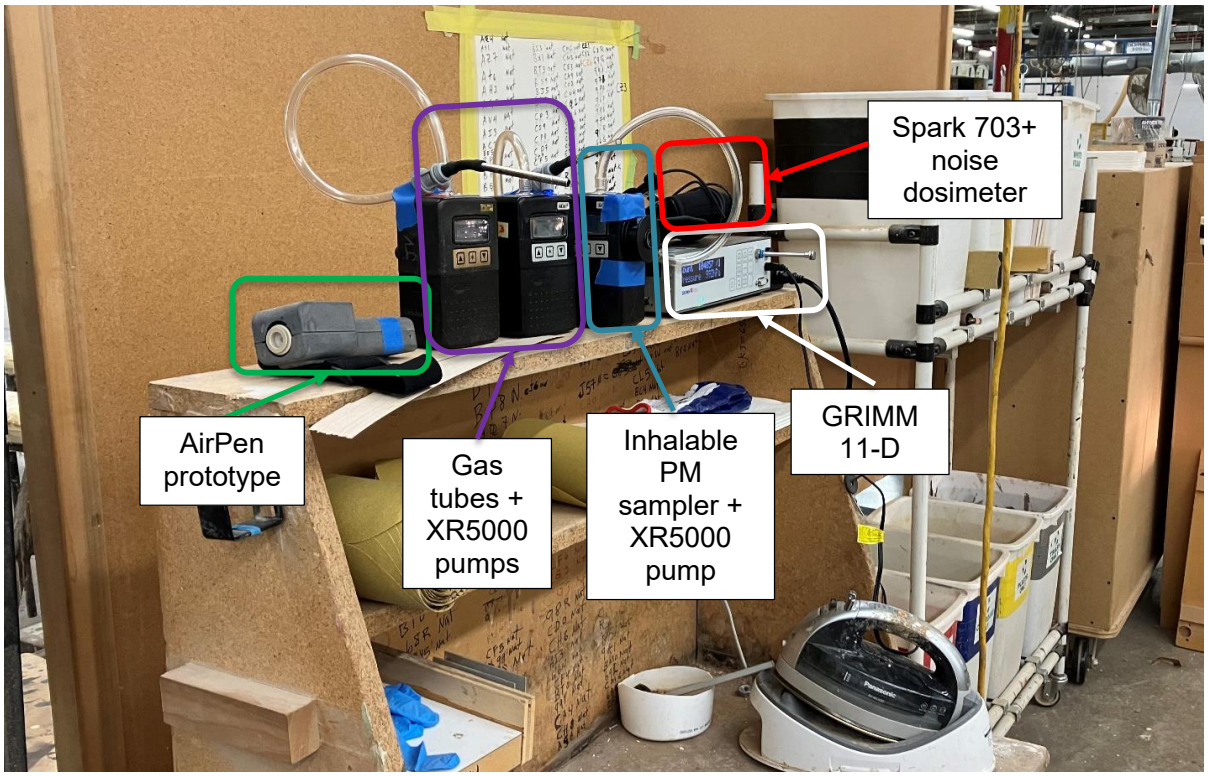

Figure B1: A prototype AirPen, conventional gas samplers, particle samplers, and a noise dosimeter colocated at a sanding workstation during the exploratory study.

1.1. PM sampling

AirPens were placed alongside PM samplers that consisted of low-cost inhalable samplers[79] connected to AirChek XR5000 pumps (SKC, Eighty Four, PA, USA). AirPen were equipped with 15-mm diameter PTFE membrane filters (PT15P-PF03, Measurement Technology Laboratories [MTL], Minneapolis, MN, USA). Low-cost inhalable samplers were equipped with 37-mm-diameter PTFE-bonded glass fiber filters (Pallflex® Emfab™). AirPens sampled air at 0.25 L min^{-1}. The XR5000 pumps sampled air at 2.0 L min^{-1}. A single portable aerosol spectrometer (11-D, GRIMM Aerosol Technik, Ainring, Germany) was deployed at a sanding workstation to measure dust mass fractions (PM_{10}, $PM_{2.5}$, PM_1, inhalable, respirable). The aerosol spectrometer data were interpreted assuming a constant particle density as a function of particle size because the dominant source of PM inside the factory was assumed to be wood cutting and sanding.

Inhalable PM concentrations in the 200 – 550 μg m^{-3} range were measured in different locations across the factory (Figure B2). The mean concentration measured by conventional samplers was 371 μg m^{-3}. At this concentration, an AirPen operating at 0.25 L min^{-1} would sample 45 μg of inhalable PM during an 8-hour shift. The time-resolved aerosol spectrometer data included peaks in PM concentrations that exceeded the mean concentrations by an order of magnitude. Besides that, inhalable mass concentrations were an order of magnitude larger than $PM_{2.5}$ concentrations (Figure B3). We decided to sample total particulates not otherwise regulated (i.e., total aerosol mass) during the full-scale deployment, in part because these preliminary results indicated that large particles were significantly more abundant than fine particles.

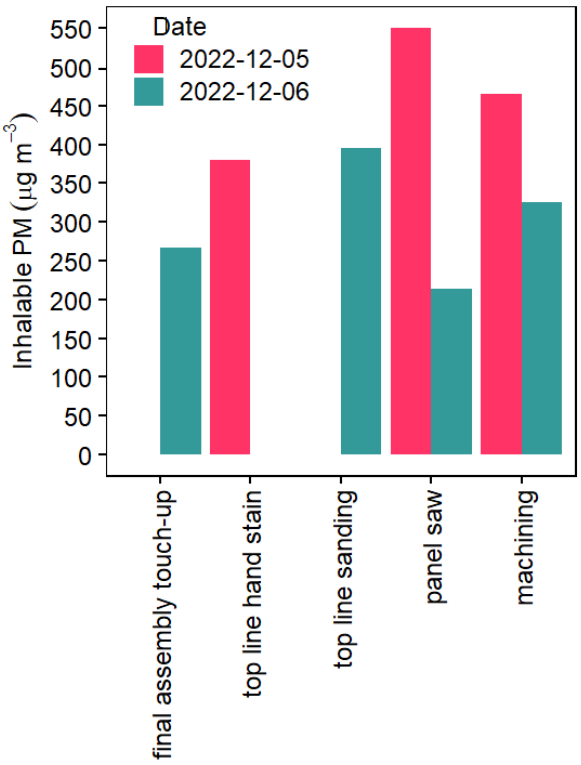

Figure B2: Inhalable PM concentrations measured in different factory lines using low-cost inhalable samplers connected to XR5000 pumps. Descriptions of production lines are available in section 2.1.

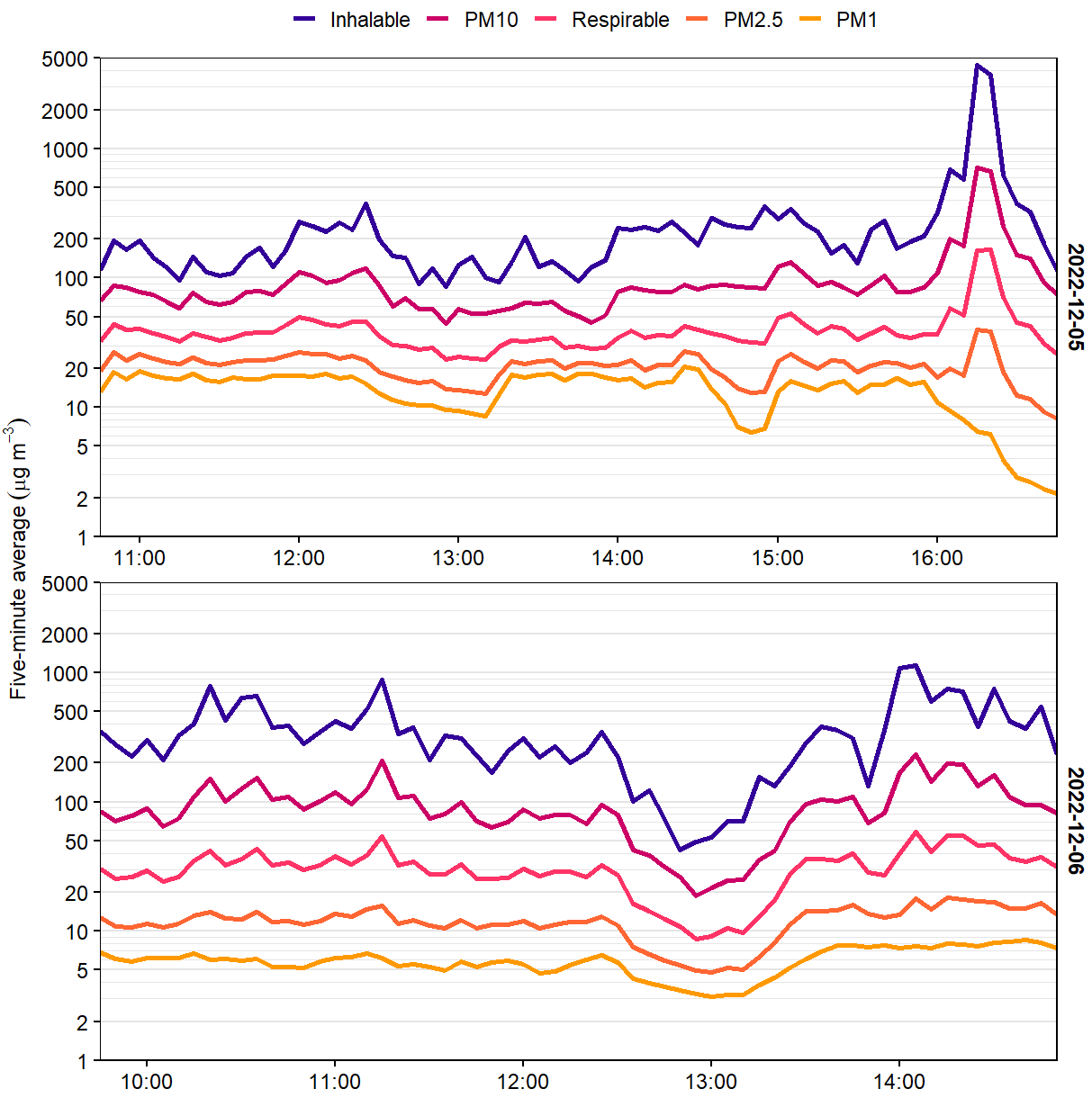

Figure B3: Dust fractions measured by the GRIMM 11-D aerosol spectrometer at a sanding station on two consecutive workdays.

1.2. Noise sampling

Spark 703+ dosimeters (Larson Davis, Depew, NY, USA) were used to collect preliminary noise data. Microphones had not been integrated into AirPens at this time. All measured noise levels were in the 65 to 88 dBA (time-weighted average) range.

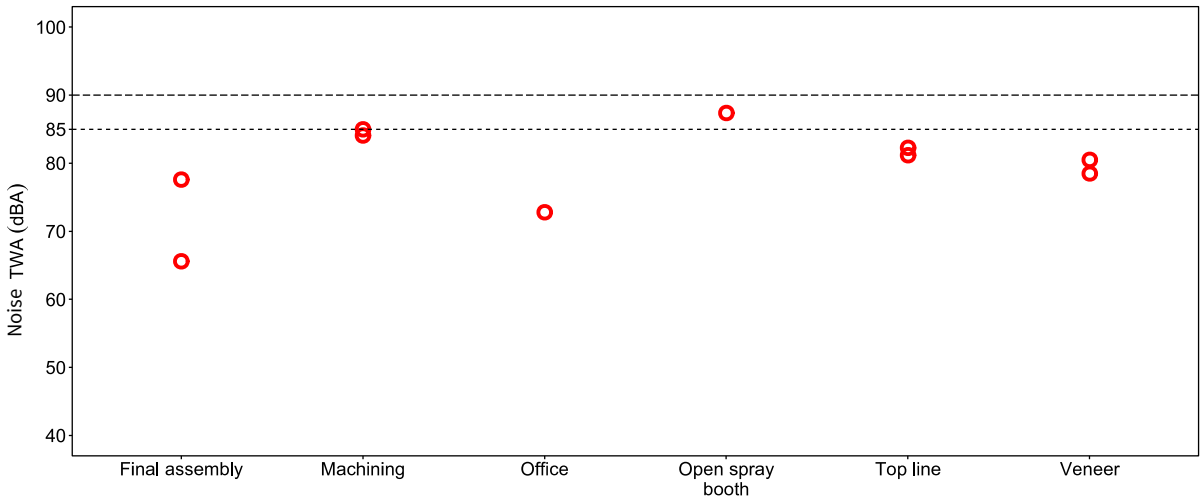

Figure B4: Time-weighted average noise levels measured with Spark 703+ dosimeters at various locations.

1.3. Carbonyl sampling

AirPens and AirChek XR5000 pumps (SKC, Eighty Four, PA, USA) were used to sample formaldehyde and other carbonyls onto ORBO™ LpDNPH glass tubes (20081-U, Sigma-Aldrich, St. Louis, MO, USA). XR5000 pumps sampled air at a flow rate of 30 mL min^{-1}. AirPens sampled air through gas tubes at 6 to 9 mL min^{-1}, depending on the AirPen unit. On each AirPen, the air flow rate through the gas tube is controlled passively by an orifice (but is not configurable) and is continuously monitored and logged. Formaldehyde concentrations ranged from 11 to 20 ppb across the different sampling locations, and acetaldehyde ranged from 1.9 to 2.8 ppb (Figure B5). Acetone had a much wider range (45 – 240 ppb) and butanone ranged from non-detect to 22 ppb. Formaldehyde measurements were above the limit of detection (LOD) when sampling ~11 ppb concentrations for 10 h using AirPens, but calculations showed that 11 ppb concentrations sampled for ~5 h samples could be below the LOD. Aldehyde concentrations derived from AirPen samples were also compared to aldehyde concentrations derived from XR5000 pump samples to assess the accuracy of the former when sampling onto LpDNPH tubes (Figure B6).

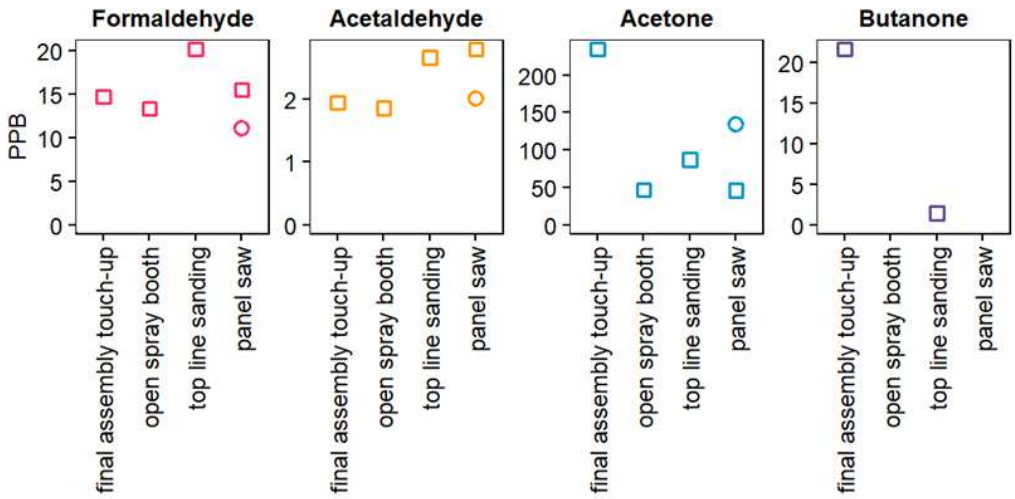

Figure B5: Carbonyl concentrations measured in different factory lines using LpDNPH glass tubes connected to XR5000 pumps. Squares correspond to samples taken during the first day (2022-12-05) and circles correspond taken during the second day (2022-12-06).

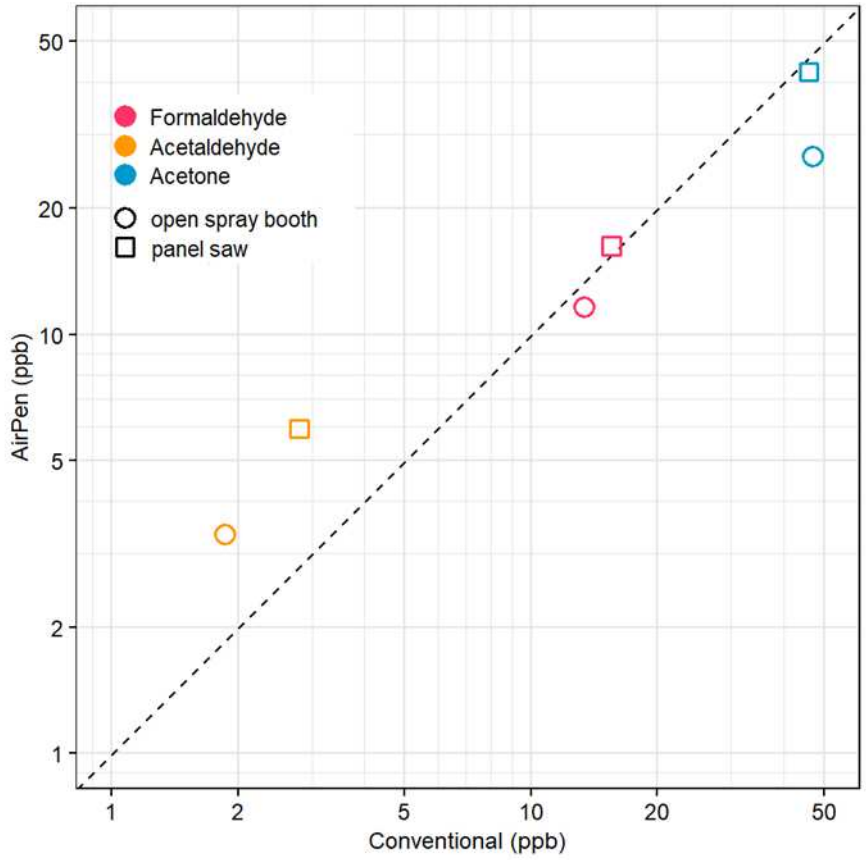

Figure B6: Comparison of carbonyl concentrations measured by colocated conventional samplers (LpDNPH glass tubes connected directly to XR5000 pumps) and AirPens. Dashed line is 1:1.

2. Study site

2.1. Production lines and departments

Employees at the facility worked in one of the following production lines or departments.

Machining (in factory): Workers operated various types of large, automated machines (e.g., CNC machining centers, double end tenoning machines) to cut, drill, and groove wooden parts. Most machines were enclosed and had dust extraction systems. Manual operations beyond loading machines or moving furniture parts were minimal.

Veneer (in factory): Workers applied thin decorative coverings to wood or fiberboard panels. Widespread use of adhesives. Extensive use of manual tools to cut, align, and press down veneer sheets. Use of large power tools or machines was minimal.

Finishing (in factory): Workers painted and stained wooden parts. Most of the work was done by hand. Parts were sanded or cleaned before and after applying paint or stain. Some spraying stations were open, and others were enclosed. Various types of paint sprayers were used.

Top line (in factory): Workers added final elements to panels before final assembly. Operations included installing multiple types of trim and decorative elements to edges, and adding functional elements (e.g., cable guides, rails, fittings). Manual and power tools were widely used, as well adhesives and fasteners. Drilling, cutting, and sanding were observed in this line.

Final assembly (in factory): Parts from different lines were put together with fasteners. Manual and power tools were widely used, but adhesives and paint were not.

Shipping (non-factory): Located at one end of the factory, next to the final assembly area. The large volume of finished furniture acted as a barrier between the loading area and the

manufacturing areas. Workers moved finished products to the loading bay and into trucks outside of the factory floor. The work performed by these workers differed significantly from the rest of the workers on the factory floor. For instance, no machining, painting, or gluing happened in this area. Therefore, these workers were classified as *non-factory*, even though they worked partially in the same building.

Admin (non-factory): Administrative workers spent most of their time in offices adjacent to the factory. However, offices and manufacturing areas were well isolated from each other (e.g., multiple sets of doors between them). A showroom was also located in the office area, so good insulation was crucial to keep it clean and quiet.

3. Study protocols

3.1. Laboratory pre-deployment

AirPens were shipped to the facility in a "ready-to-use" state. Prior to shipping the AirPens to the facility for deployment, sampling media (i.e., a PTFE filter and LpDNPH glass tube) were installed in each AirPen in a laboratory at Colorado State University. All sampling media were handled with nitrile gloves and clean tweezers. Ozone scrubbers were not used upstream of the LpDNPH glass tubes given that we were sampling indoors during winter (December 2023). Ready-to-use samplers were individually vacuum sealed and packed in hard cases with custom foam dividers (Figure B7).

Figure B7: "Ready-to-use" AirPens vacuum sealed individually and packed in a hard case with foam dividers. Samplers were shipped in this setup from the laboratory to the field.

3.2. Field pre-deployment

Two days before the deployment, in a hotel near the facility, AirPens were taken out of the hard cases, vacuum sealed bags were opened, and batteries were charged overnight. Field blank AirPens were also unpacked and positioned in the same area where functional AirPens were charging. Once the batteries were fully charged, we turned on the devices and checked their status through the iOS app. A physical inspection revealed that many of the LpDNPH glass tubes had broken during shipment. Those AirPens had to be disassembled so that the broken tubes could be removed; then, those AirPens had to be cleaned, and the gas sampling channel within each affected AirPen had to be capped.

Below is an example of a broken LpDNPH glass tube in an AirPen. The DNPH-coated silica beads and pieces of broken glass were thoroughly cleaned out of each affected AirPen, and the

inlet and outlet of the gas sampling channel on the AirPen were capped, before each AirPen was reassembled.

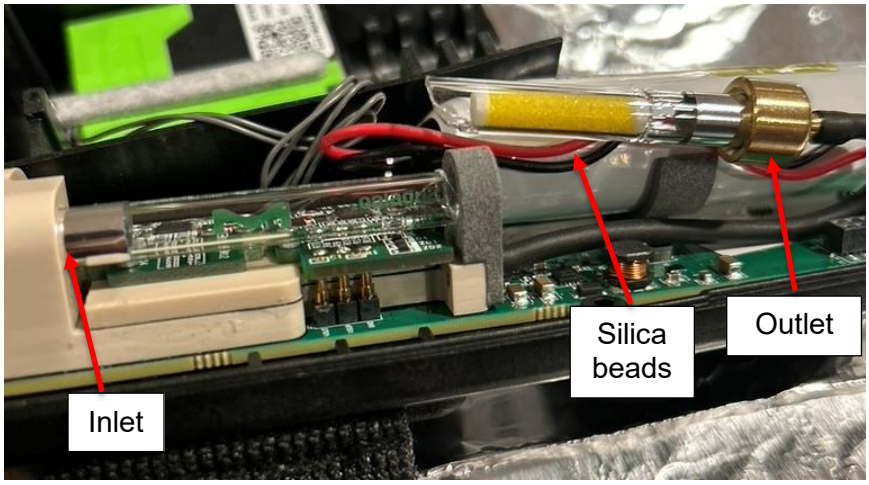

Figure B8: An AirPen with a broken glass LpDNPH tube.

After charging the batteries, and fixing units with broken LpDNPH glass tubes, AirPens were packed in their hard cases to be transported from the hotel to the factory.

Figure B9: Ready-to-use AirPens in a hard case. Samplers were stored in this setup overnight before the deployment.

3.3. Participant recruitment

Recruitment efforts for the project started approximately one month before the AirPen deployment and included flyers, emails, and in-person visits. To receive an AirPen during the sampling campaign, each worker had to complete a pre-sampling intake survey. Most participants were recruited and completed the intake survey during a field team visit that occurred two weeks before the exposure sampling day, but some new participants were recruited on the sampling day.

On the day of deployment, the study team had three tables set up at various points of the facility to distribute AirPens. Consent forms and pre-deployment surveys were available at each table. Field team members provided information about the study and guided newly-recruited participants through the consent and intake survey processes.

Figure B10: A member of the field team recruiting a participant.

3.4. Sampling and media handling

The field sampling team was comprised of two members of a "technology team" (including an AirPen software/hardware developer) and three members of a "communications/social science

team.” Only AirPens were deployed on the facility-wide sampling day, so the process consisted of fitting a sampler around each participant’s arm and turning it on. After sampling during the work shift, AirPens were kept in the hard shipping cases overnight, not in sealed plastic bags. The next day, filters and gas tubes were removed from the AirPens and transferred to individual sealed containers. This process was carried out in a hotel lobby. All work surfaces were cleaned and covered with aluminum foil to minimize sample contamination; additionally, clean nitrile gloves and tweezers were used to handle sampling media.

Figure B11: Three participants wearing AirPens around their arms.

4. Sample analysis

4.1. Total particulates not otherwise regulated

Each PTFE membrane filter was pre- and post-weighed in triplicate using an automated system that consisted of a microbalance with 1-µg resolution (XS3DU, Mettler Toledo, Columbus, OH, USA) inside a climate-controlled enclosure.[83] If the first three replicate weights had a range greater than 5 µg, the filter was re-weighed until three consecutive weights with a range ≤ 5 µg were obtained. The mass on each filter was calculated as the difference between the average of the post-weights and the average of the pre-weights, minus the average blank mass. To obtain the total PM concentration in air, the mass on the filter was divided by the total air volume sampled by the AirPen.

The PM limit of detection (LOD; µg) was calculated as shown in Equation B1:

$$LOD = \bar{x}_{blank} + 3s_{blank} \qquad (B1)$$

where $\bar{x}_{blank}$ was the mean and s_{blank} was the standard deviation of the changes in the masses of the blank filters collected during the field campaign (µg). Blank filters were pre-weighed, installed in non-functional (but otherwise identical) AirPens, shipped to the field, and post-weighed along with sample filters, but no air was sampled through blank filters. The PM LOD was 10.4 µg (calculated from 13 blanks with a mean change in mass of 3.7 µg). The AirPen PM sample volumetric flow rate (250 mL min^{-1}) corresponded to a sampled volume of 0.12 m^3 for an 8-hour sample. Thus, the 8-hour PM LOD was 86.7 µg m^{-3}.

4.2. Formaldehyde

The contents of each tube were extracted using 3 mL of a mixture of 70% acetonitrile and 30% deionized water drawn through the cartridge using a vacuum manifold. The extract was collected in a 5 mL volumetric flask. Once extraction was complete, the volumetric flask was filled to a final extraction volume with the 70% acetonitrile:30% deionized water. Each extract was allowed to equilibrate for 20 min following the instrument manufacturer's instructions and then analyzed immediately on an Agilent 1260 Infinity HPLC (Santa Clara, CA, USA) equipped with a dual-channel pump and UV detector set to monitor light intensity at a wavelength of 360 nm. The separation was conducted on a Waters Nova-Pak (Milford, Massachusetts, USA) C-18 4-µm column (3.9 × 150 mm). The eluents were 60% deionized water:30% acetonitrile:10% tetrahydrofuran (A) and 40% water:60% acetonitrile (B). The complete run time was 32 minutes and included 2 steps. For the first 25 minutes, a linear gradient from 100% A to 100% B was conducted. Then, for the final 5 minutes, a re-equilibration step was done to return to the starting conditions of 100% A. The flow rate during each step was 1.5 mL min^{-1}. An injection volume of 10 µL was used for each sample.

The HPLC reported the peak areas for several carbonyls, including formaldehyde, in absorbance units. We subtracted the mean peak area of the blanks from the value reported by the instrument to account for bias introduced during handling and shipping. Peak areas were transformed to equivalent formaldehyde mass using a calibration curve. The calibration curve was obtained using a mixed carbonyl standard solution in DNPH (CRM47285, Sigma-Aldrich, St. Louis, MO, USA). The formaldehyde mass was divided by the volume of air sampled through the LpDNPH tube on each AirPen to obtain the average concentration in air.

The formaldehyde LOD was calculated using Equation B1 where $\bar{x}_{blank}$ was the mean and s_{blank} was the standard deviation of the peak areas for the blank LpDNPH glass tubes collected during the field campaign. Blank tubes were installed in non-functional (but otherwise identical) AirPens, shipped to the field, and analyzed along with sample tubes, but no air was sampled through blank tubes. Transformed to mass, the formaldehyde LOD was 3.1 ng (calculated from 4 blanks). Because the volumetric flow rate through the gas tubes was not actively controlled by the AirPen, the volume of air sampled through the LpDNPH tube varied between samples. The formaldehyde LOD was equivalent to concentrations of 17.5, 10.1, and 6.5 μg m^{-3} for the minimum, median, and maximum sampled volumes, respectively. These LODs were equivalent to 14.3, 8.2, and 5.3 ppb at 1 atm and 25 °C (the study site was near sea level and indoors).

5. Flow control analysis

During the sample validation process, flow rate and pump voltage data recorded by sensors on AirPens were used to identify flow control anomalies. The first and last 5 minutes of data from each sample were disregarded to account for the initial flow control stabilization process as well as pre- and post-sampling manipulation. Individual (i.e., 30-s) observations for which the flow rate recorded for the filter sample deviated from the target (0.25 L min^{-1}) by more than 10% were flagged as point anomalies. For pump voltage, a 10-minute rolling mean (still at 30-s resolution) was used to catch only longer-term trends. Changes of the pump voltage rolling mean larger than 2.5 V h^{-1} were flagged as point anomalies. Samples with 5 or more minutes of anomalous data (i.e., ≥ 10 point anomalies) were flagged as potentially anomalous samples. The final decision of whether a sample was considered anomalous was based on a visual assessment and comparison with known faulty behavior (examples below).

5.1. Example of adequate flow control (non-anomalous reference)

The example below corresponds to one of the samples collected during the facility-wide campaign and represents a good reference of how AirPens perform when everything works as intended. The anomaly detection algorithm did not flag any of the timestamps as anomalous. Visually, both the pump voltage and flow rate remained steady throughout the day (Figure B12).

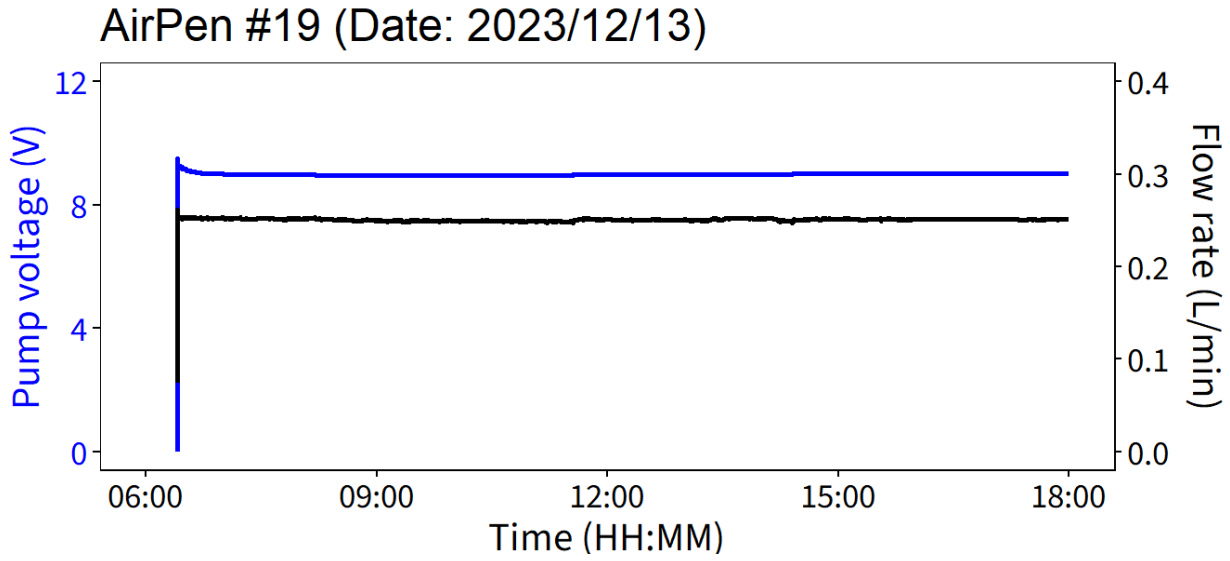

Figure B12: Example of a sample with flawless flow control. Blue line shows the pump voltage (V), and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}).

5.2. Example of leaking AirPen (anomalous reference)

The example below was recorded during a laboratory experiment carried out approximately six months before the deployment at the furniture factory. Upon inspection of the experiment results, a bias was identified relative to the reference instrument. The unit was checked under vacuum and a leak was found. The flow control anomaly detection algorithm described in the main document was applied to data from this laboratory sample to demonstrate the algorithm's ability to detect leaks. Figure B13 shows the sustained increase in pump voltage that characterizes a slow leak. The figure also shows the flow rate dropping below the acceptable range (0.25 L min^{-1}

± 10%) once the pump voltage cannot increase any further to compensate for the leak. Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm. However, a visual assessment of the overall trend is necessary to understand the device's behavior.

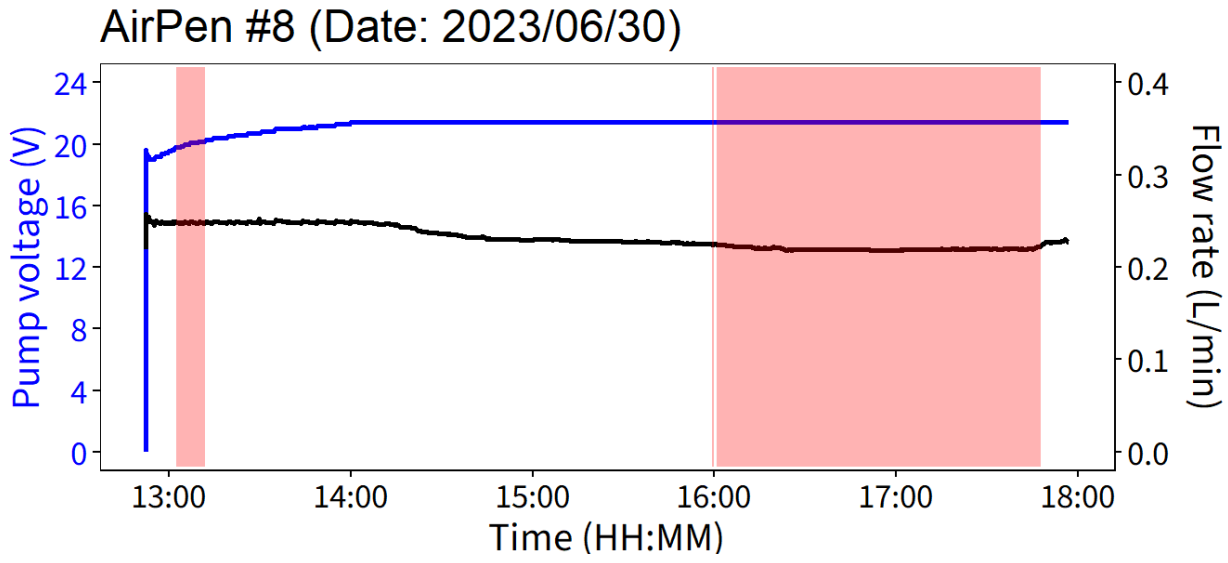

Figure B13: Example of a sample from a leaking AirPen. Blue line shows the pump voltage (V), and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm.

5.3. Sample correctly flagged as anomalous

The following sample from the facility-wide personal sampling campaign was flagged as potentially anomalous and visual inspection confirmed the existence of a leak. The sudden flow rate drops accompanied by abrupt increases in pump voltage are consistent with an AirPen trying to compensate for a leak. During the segments of relatively steady flow that were recorded throughout the day, the pump voltage ranged from 9V to 13V throughout the day, which is an excessively large range.

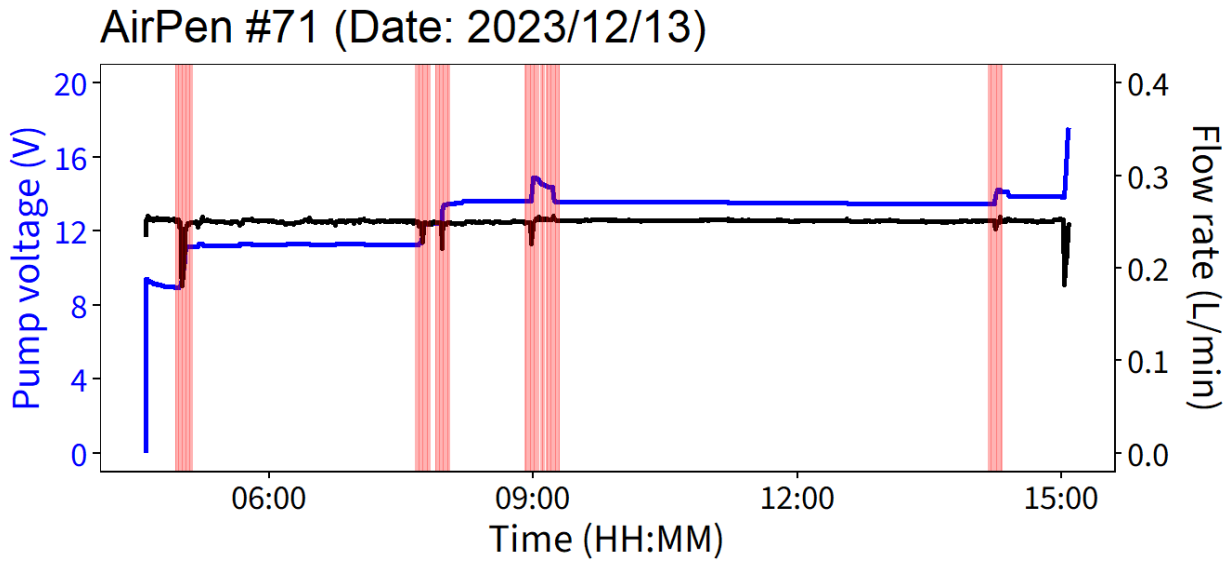

Figure B14: Example of a sample from a leaking AirPen. Blue line shows the pump voltage (V), and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm. This whole sample was designated as anomalous and removed from the dataset during the sample validation process.

5.4. Samples flagged as potentially anomalous, deemed valid upon visual assessment

Samples in this subsection were flagged by the algorithm as potentially anomalous because they included 5 or more minutes of anomalous flow control data, but visual inspection did not identify a problematic trend. Flagged segments were short and most changes in voltage or flow rate were reverted after a few minutes. These acute events were probably caused by external stimuli. These samples were *not* removed from the dataset during the sample validation process.

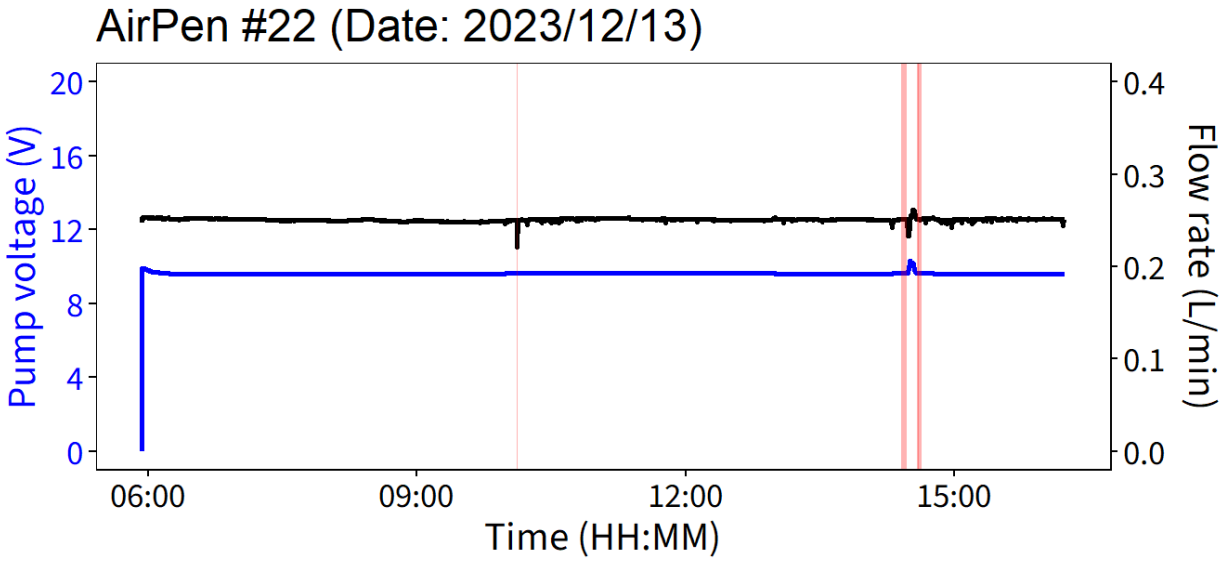

Figure B15: Sample flagged as potentially anomalous by the algorithm but deemed valid upon visual inspection. Blue line shows the pump voltage (V), and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm.

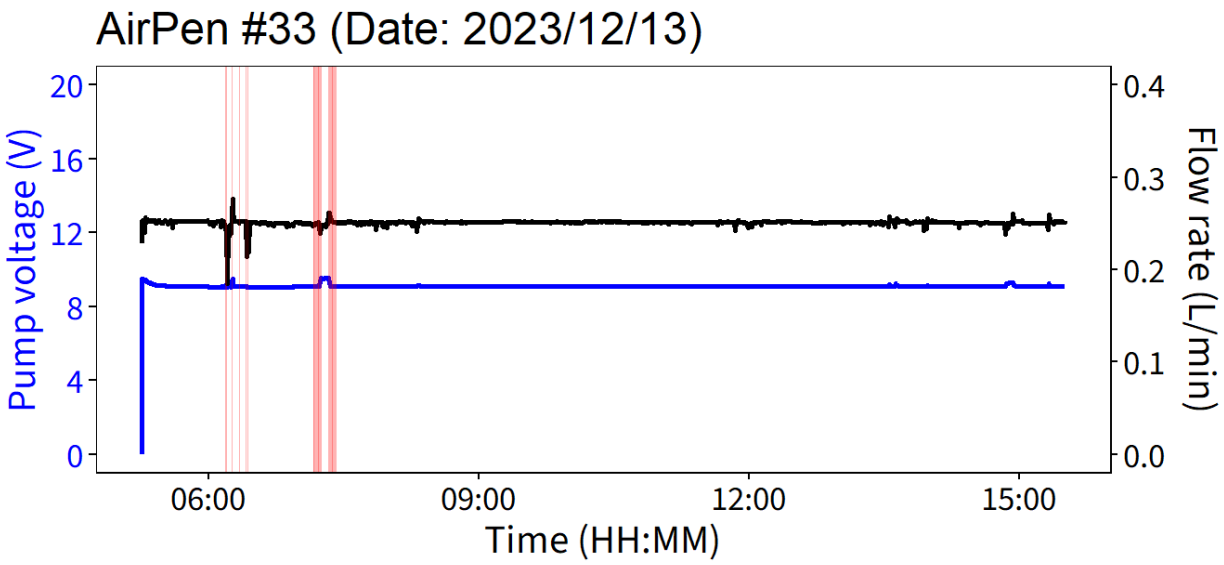

Figure B16: Sample flagged as potentially anomalous by the algorithm but deemed valid upon visual inspection. Blue line shows the pump voltage (V), and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm.

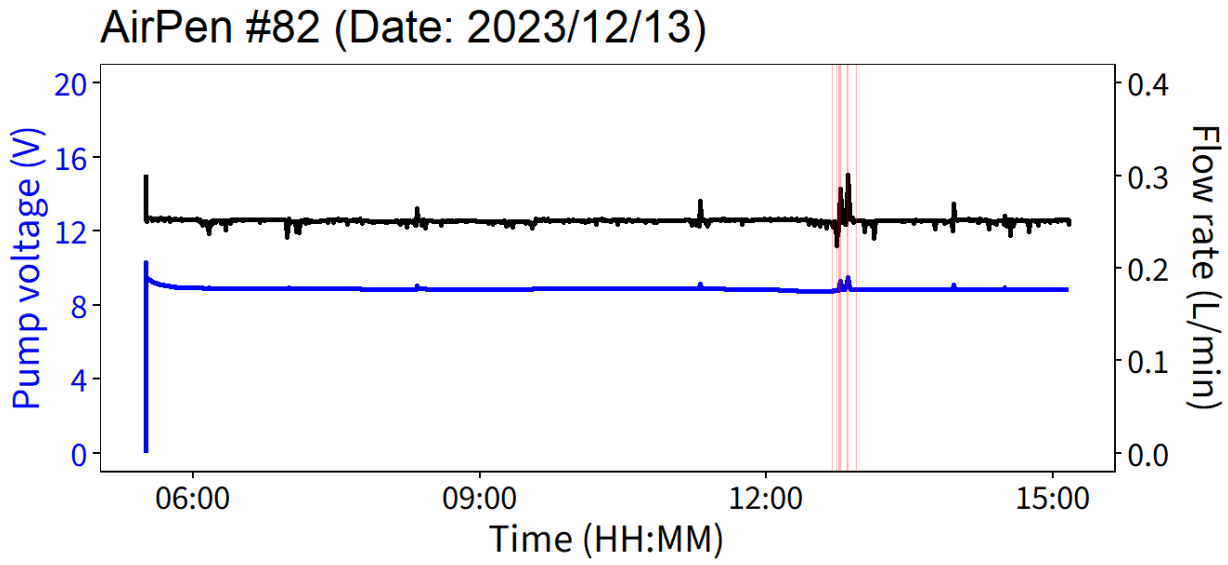

Figure B17: Sample flagged as potentially anomalous by the algorithm but deemed valid upon visual inspection. Blue line shows the pump voltage (V) and black line shows the flow rate through the AirPen inlet and PM sample filter (L min^{-1}). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm.

6. Wearing compliance analysis

During the sample validation process, AirPen accelerometer data were analyzed to confirm that each AirPen was worn by the worker for the full duration of the work shift and therefore captured the worker's true personal exposures. AirPen accelerometer data were processed on-board at 1 Hz frequency, and the following variables were logged in 30-second intervals: step count, 3-axis mean accelerations, 3-axis minimum and maximum accelerations, and 3-axis acceleration variances. Acceleration entropy, variance, and frequency can all be used for motion detection and activity recognition.[140,141] Out of these three metrics, the AirPen only calculated variance on-board because variance has the lowest computational complexity. Accelerometer data were time-averaged to 5-minute intervals. Z-scores were then calculated for the acceleration variances of each axis. Observations with a z-score of less than negative three (i.e., three standard deviations steadier than the mean) for any of the three axes were flagged as point anomalies. This approach was possible because the field team observed visually during the

sample day that AirPens were largely worn by workers as requested, so AirPens that were not worn could be considered outliers. The first and last five minutes were not considered for the reasons described in the previous section. Samples with more than 20% of anomalous data (i.e., 96 minutes of total anomalous data for an 8-hour sample; not necessarily consecutive) were flagged as potentially anomalous samples. Acceleration and step count graphs of potentially anomalous samples were visually checked to ensure that the flagged events were consistent with the worker not wearing the AirPen and not triggered by instrument malfunction (e.g., an unresponsive/damaged accelerometer would be flagged by the algorithm but would not provide any real information).

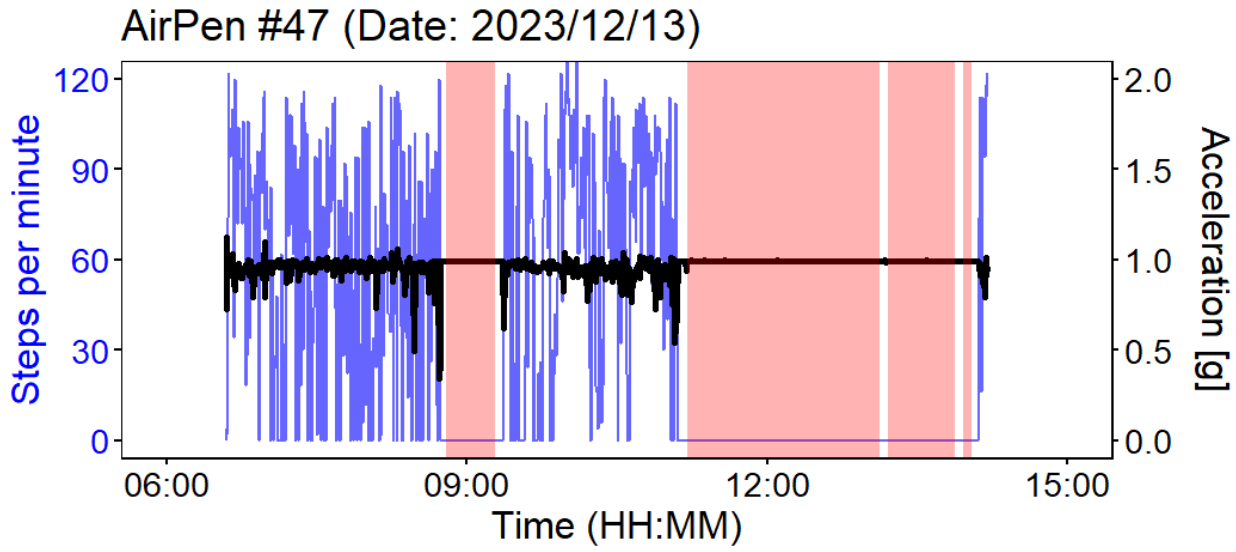

Figure B18: Accelerometer data from an AirPen sample that was designated as anomalous and removed from the dataset during the sample validation process because there were long periods of time during the work shift when the AirPen was not worn.

An example sample that was excluded because the AirPen was not worn for the full work shift is shown in Figure B18. Accelerometer data from this sample indicated that there were more than three hours total sample time during which the AirPen was not worn. Areas highlighted in red were flagged as abnormally steady events by the algorithm described in the main document. The

figure shows that the magnitude of the combined accelerations was remarkably steady and practically equivalent to gravity (i.e., 1 g) during those events, so it is fair to assume that gravity was the only force acting on the devices. Step count was zero during these anomalous events.

7. Extended formaldehyde results

7.1. Fitted distributions

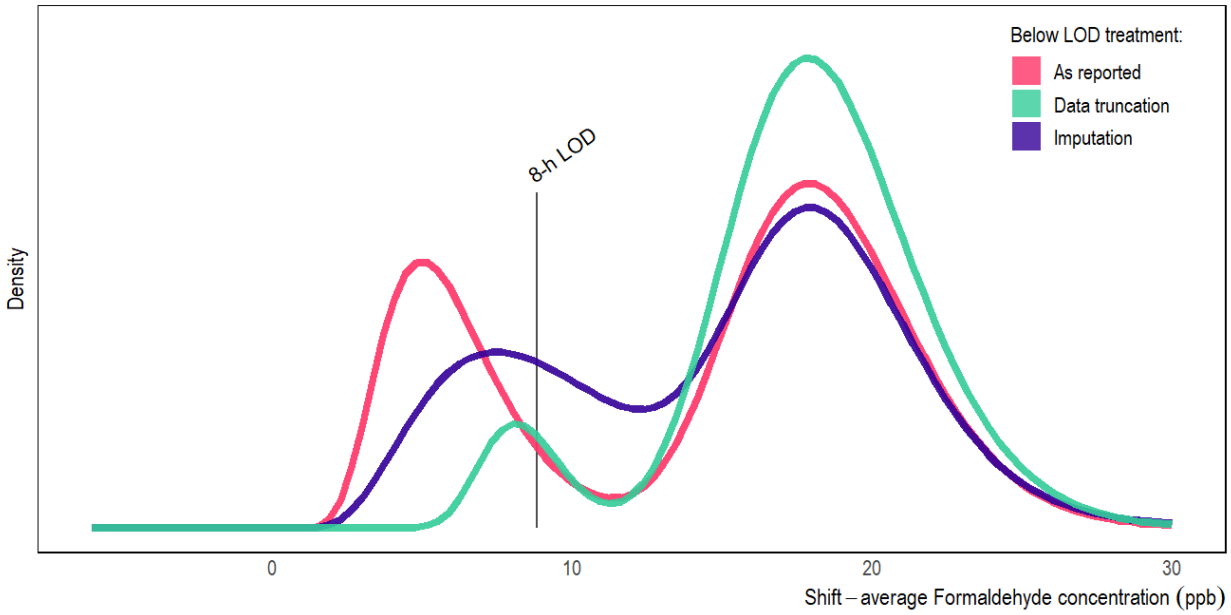

Figure B19: Formaldehyde exposure distributions shown as fitted bimodal lognormal probability density functions. Colors represent the left-censored data treatment.

7.2. Summary statistics

Parameters for the weighted mixture of two lognormal (i.e., bimodal lognormal) fits (i.e., μ, σ, ϕ) shown in Figure B19 were estimated using the *VGAM* package in R.[142] The rest of the statistics presented below were calculated using empirical measurements and not the distribution fits.

Table B1: Formaldehyde summary statistics. Distribution parameters, central tendency, dispersion, and upper percentile estimates. Statistics are disaggregated by left-censored data treatment.

Formaldehyde [ppb]			
	Below LOD treatment		
Parameter	*Truncation*	*Imputation*	*Data as reported*
Estimated from fitted distribution			
1st Meanlog μ_1 [log(ppb)]	2.120	2.246	1.744
1st SDlog σ_1 [log(ppb)]	0.165	0.485	0.368
2nd Meanlog μ_2 [log(ppb)]	2.909	2.922	2.911
2nd SDlog σ_2 [log(ppb)]	0.166	0.159	0.165
Mix. Coef. ϕ [unitless]	0.091	0.462	0.339
Estimated from raw data			
Sample size	22	31	31
GM	17.08	13.60	12.38
GSD [unitless]	1.333	1.638	1.855
95th pct.	23.68	23.15	23.15
Minimum	7.09	2.80	-4.38
Maximum	25.57	25.57	25.57
Arithmetic mean	17.68	14.99	13.79

8. Extended similar exposure groups (SEG) results

8.1. Particulate matter

The particulate matter exposure distribution was clearly lognormal, so all the following tests were applied to log-transformed PM concentrations to compare exposures between workers from different departments within the facility and different production lines within the factory.

Welch's two-sample t-test was used to determine whether mean exposure to PM was significantly different between factor workers and non-factory workers. Non-factory workers

were those who worked in shipping or administrative roles. The output from this test, shown in Figure B20, indicated that the mean PM exposure was different between factory and non-factory workers at a significance level of 0.05.

```
##
##  Welch Two Sample t-test
##
## data:  log(ug_m3) by task
## t = 4.6123, df = 8.5652, p-value = 0.001444
## alternative hypothesis: true difference in means between group Factory and group
Shipping/Admin is not equal to 0
## 95 percent confidence interval:
##  0.413643 1.222221
## sample estimates:
##        mean in group Factory mean in group Shipping/Admin
##                     5.623288                     4.805355
```

Figure B20: Output of the Welch t-test for log-transformed PM exposure. Groups compared in this test were factory and non-factory (i.e., shipping and admin) workers.

A one-way analysis of variance (ANOVA) test was used to evaluate whether mean PM exposures were significantly different between production lines within the factory. The output from this test, shown in Figure B21, indicated that PM exposures were not significantly different between production lines.

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## task         4  1.205  0.3014    0.92  0.463
## Residuals   38 12.453  0.3277
```

Figure B21: Output of the one-way ANOVA test for log-transformed PM exposure. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

Tukey multiple pairwise comparisons were used to evaluate whether mean PM exposures were significantly different between any two production lines within the factory. The output from this test, shown in Figure B22, indicated that PM exposures were not significantly different between any two production lines.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = log(ug_m3) ~ task, data = subset(df_pm_anova, task != "Unknow
n" & task != "Admin" & task != "Shipping"))
##
## $task
##                                 diff        lwr       upr     p adj
## Top Line-Finishing          0.01418735 -0.7338921 0.7622668 0.9999979
## Veneer-Finishing            0.03342441 -0.8517151 0.9185639 0.9999676
## Machining-Finishing        -0.35485428 -1.1512457 0.4415371 0.7073327
## Final Assembly-Finishing   -0.29195853 -1.1114385 0.5275214 0.8445540
## Veneer-Top Line             0.01923706 -0.8002429 0.8387170 0.9999951
## Machining-Top Line         -0.36904163 -1.0917550 0.3536718 0.5926586
## Final Assembly-Top Line    -0.30614588 -1.0542253 0.4419335 0.7671558
## Machining-Veneer           -0.38827869 -1.2520864 0.4755290 0.7006384
## Final Assembly-Veneer      -0.32538294 -1.2105225 0.5597566 0.8291648
## Final Assembly-Machining    0.06289575 -0.7334956 0.8592871 0.9993928
```

Figure B22: Tukey multiple pairwise comparison for log-transformed PM exposure. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

The residuals plot in Figure B23 was used to visually identify variance patterns that would be a sign of a violation of the homoscedasticity assumption (e.g., funnel or cornucopia shapes) required for the ANOVA test. No evidence of a violation of the assumption was found for the log-transformed PM exposure data.

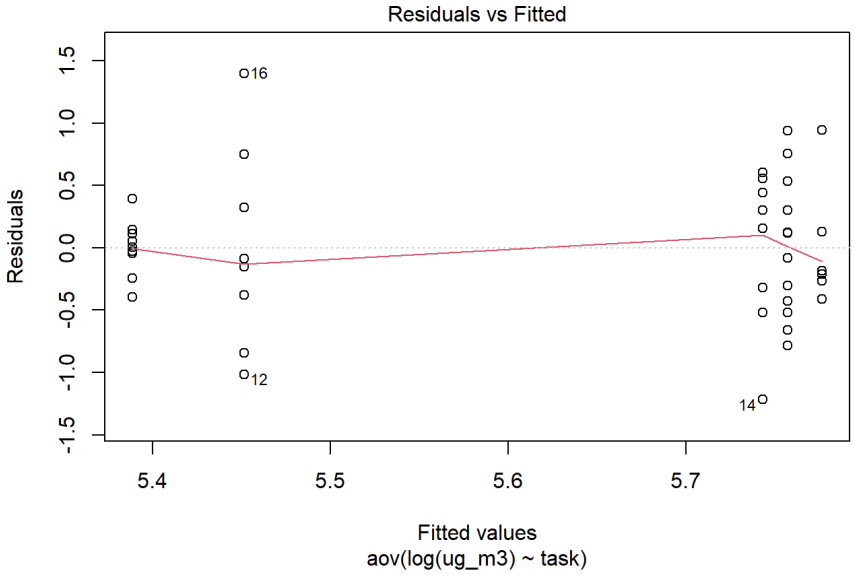

Figure B23: Residuals plot for log-transformed PM exposure.

Levene's test is a probabilistic approach for detecting violations of the homoscedasticity violation. This test was used to confirm the visual assessment described in the previous subsection (i.e., the residuals plot approach). The null hypothesis of Levene's test is that the variances are homogeneous. The *p*-value of 0.1497 shown in Figure B24 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the homogeneity of variance assumption for the log-transformed PM exposure data.

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  1.7964 0.1497
##       38
```

Figure B24: Output of Levene's test for log-transformed PM exposure.

The Q-Q plot in Figure B25 was used to visually assess the normality of the log-transformed PM exposure data. No major deviations from the line of normality (i.e., the 1:1 line) were observed. No violations of the normality assumption were found.

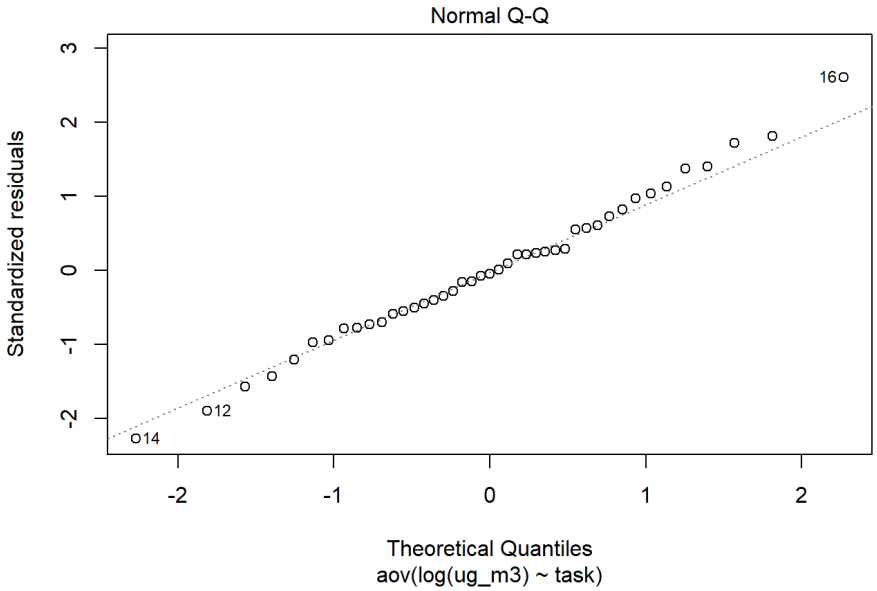

Figure B25: Q-Q plot for log-transformed PM exposure.

The Shapiro-Wilk test is a probabilistic approach for detecting violations of the normality assumption required for the ANOVA test. This test was used to confirm the visual assessment described in the previous subsection (i.e., the Q-Q plot approach). The null hypothesis of the Shapiro-Wilk test is that data comes from a normal distribution. The *p*-value of 0.9955 shown in Figure B26 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the normality assumption for the log-transformed PM exposure data.

```
##
##  Shapiro-Wilk normality test
##
## data:  aov_residuals
## W = 0.99306, p-value = 0.9955
```

Figure B26: Output of Shapiro-Wilk's normality test for log-transformed PM exposure.

8.2. Noise

The noise exposure distribution was close to normal, so all the following tests were applied to the TWA noise measurements to compare exposures between workers from different departments within the facility and different production lines within the factory.

Welch's two-sample t-test was used to determine whether mean exposure to noise was significantly different between factor workers and non-factory workers. Non-factory workers were those who worked in shipping or administrative roles. The output from this test, shown in Figure B27, indicated that the mean noise exposure was different between factory and non-factory workers at a significance level of 0.05.

```
##
##  Welch Two Sample t-test
##
## data:  TWA_dBA by task
## t = 8.5329, df = 11.151, p-value = 3.208e-06
## alternative hypothesis: true difference in means between group Factory and group
Shipping/Admin is not equal to 0
## 95 percent confidence interval:
##   6.385092 10.814213
## sample estimates:
##        mean in group Factory mean in group Shipping/Admin
##                     78.78709                     70.18743
```

Figure B27: Output of the Welch t-test for noise. Groups compared in this test were factory and non-factory (i.e., shipping and admin) workers.

A one-way analysis of variance (ANOVA) test was used to evaluate whether mean noise exposures were significantly different between production lines within the factory. The output from this test, shown in Figure B28, indicated that noise exposures were not significantly different between production lines.

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## task         4   51.8   12.94   1.496  0.223
## Residuals   37  320.0    8.65
```

Figure B28: Output of the one-way ANOVA test for noise. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

Tukey multiple pairwise comparisons were used to evaluate whether mean noise exposures were significantly different between any two production lines within the factory. The output from this test, shown in Figure B29, indicated that noise exposures were not significantly different between any two production lines.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = TWA_dBA ~ task, data = subset(df_noise_anova, task != "Unknow
n" & task != "Admin" & task != "Shipping"))
##
## $task
##                                 diff       lwr       upr     p adj
## Finishing-Machining      -1.8873730 -6.042473 2.2677266 0.6915106
## Top Line-Machining       -2.2276439 -6.382743 1.9274557 0.5458665
## Veneer-Machining         -3.0745580 -7.581396 1.4322803 0.3072162
## Final Assembly-Machining -3.4081395 -7.771867 0.9555878 0.1882203
## Top Line-Finishing       -0.3402709 -4.110962 3.4304205 0.9989665
## Veneer-Finishing         -1.1871850 -5.342285 2.9679146 0.9230587
## Final Assembly-Finishing -1.5207666 -5.520189 2.4786556 0.8104470
## Veneer-Top Line          -0.8469141 -5.002014 3.3081855 0.9765882
## Final Assembly-Top Line  -1.1804957 -5.179918 2.8189265 0.9142384
## Final Assembly-Veneer    -0.3335816 -4.697309 4.0301458 0.9994626
```

Figure B29: Tukey multiple pairwise comparison for noise. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

The residuals plot in Figure B30 was used to visually identify variance patterns that would be a sign of a violation of the homoscedasticity assumption (e.g., funnel or cornucopia shapes) required for the ANOVA test. No evidence of a violation of the assumption was found for the noise exposure data.

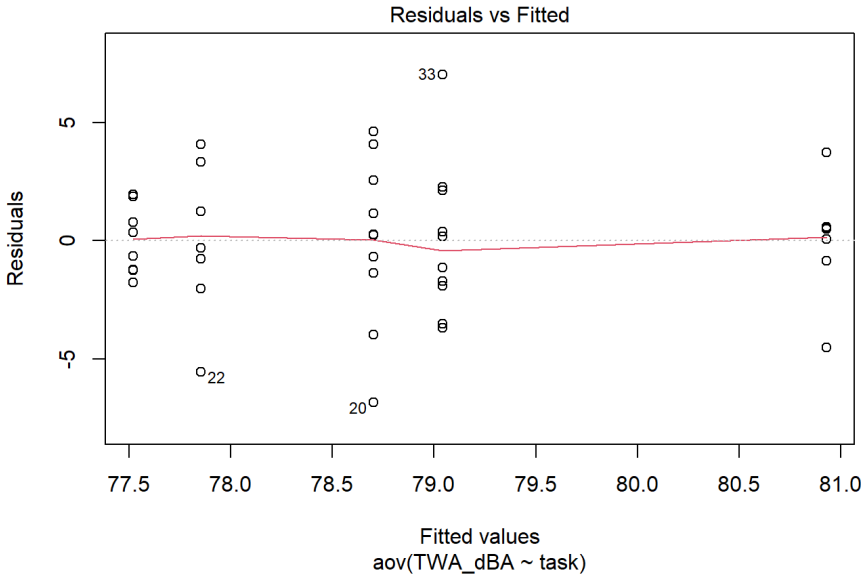

Figure B30: Residuals plot for noise.

Levene's test is a probabilistic approach for detecting violations of the homoscedasticity violation. This test was used to confirm the visual assessment described in the previous subsection (i.e., the residuals plot approach). The null hypothesis of Levene's test is that the variances are homogeneous. The *p*-value of 0.5017 shown in Figure B31 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the homogeneity of variance assumption for the noise exposure data.

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  0.8518 0.5017
##       37
```

Figure B31: Output of Levene's test for noise.

The Q-Q plot in Figure B32 was used to visually assess the normality of the noise exposure data. No major deviations from the line of normality (i.e., the 1:1 line) were observed. No violations of the normality assumption were found.

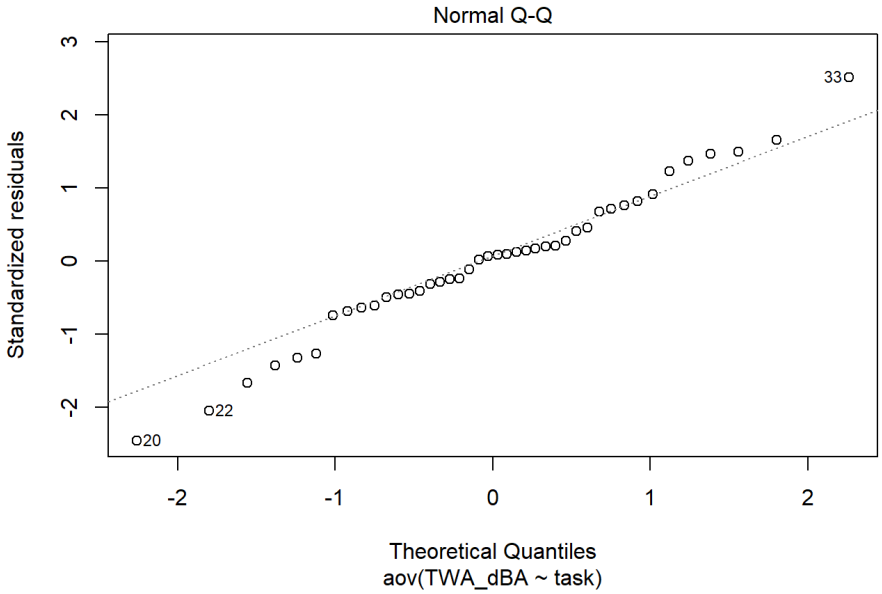

Figure B32: Q-Q plot for noise.

The Shapiro-Wilk test is a probabilistic approach for detecting violations of the normality assumption required for the ANOVA test. This test was used to confirm the visual assessment described in the previous subsection (i.e., the Q-Q plot approach). The null hypothesis of the Shapiro-Wilk test is that data comes from a normal distribution. The *p*-value of 0.8597 shown in Figure B33 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the normality assumption for the noise exposure data.

```
##
##  Shapiro-Wilk normality test
##
## data:  aov_residuals
## W = 0.98539, p-value = 0.8597
```

Figure B33: Output of Shapiro-Wilk's normality test for noise.

8.3. Formaldehyde

Even though we found the formaldehyde exposure distribution to be best represented by a bimodal lognormal shape, we applied all the following tests to untransformed formaldehyde concentrations to compare exposures between workers from different departments within the facility and different production lines within the factory. Welch's test and ANOVA assume normality and homoscedasticity but are robust to moderate deviations. The diagnostic tests in this section did not find any significant violations of the assumptions.

Welch's two-sample t-test was used to determine whether mean exposure to formaldehyde was significantly different between factor workers and non-factory workers. Non-factory workers were those who worked in shipping or administrative roles. The output from this test, shown in

Figure B34, indicated that the mean formaldehyde exposure was different between factory and non-factory workers at a significance level of 0.05.

```
##
##  Welch Two Sample t-test
##
## data:  Formaldehyde by task
## t = 3.0768, df = 4.2715, p-value = 0.03395
## alternative hypothesis: true difference in means between group Factory and group
Shipping/Admin is not equal to 0
## 95 percent confidence interval:
##   0.6865652 10.7761496
## sample estimates:
##        mean in group Factory mean in group Shipping/Admin
##                     18.41391                     12.68256
```

Figure B34: Output of the Welch t-test for formaldehyde. Groups compared in this test were factory and non-factory (i.e., shipping and admin) workers.

A one-way analysis of variance (ANOVA) test was used to evaluate whether mean formaldehyde exposures were significantly different between production lines within the factory. The output from this test, shown in Figure B35, indicated that formaldehyde exposures were not significantly different between production lines.

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## task         4  56.48   14.12   0.722  0.595
## Residuals   11 215.16   19.56
```

Figure B35: Output of the one-way ANOVA test for formaldehyde. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

Tukey multiple pairwise comparisons were used to evaluate whether mean formaldehyde exposures were significantly different between any two production lines within the factory. The output from this test, shown in Figure B36, indicated that formaldehyde exposures were not significantly different between any two production lines.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Formaldehyde ~ task, data = subset(df_formaldehyde_anova, tas
k != "Unknown" & task != "Admin" & task != "Shipping"))
##
## $task
##                                 diff       lwr       upr     p adj
## Top Line-Veneer            0.1976862 -11.76894 12.164308 0.9999978
## Finishing-Veneer          -3.3536714 -17.65652 10.949177 0.9373880
## Final Assembly-Veneer     -3.6742637 -15.64089  8.292358 0.8531567
## Machining-Veneer          -3.8240454 -18.12689 10.478803 0.9039382
## Finishing-Top Line        -3.5513576 -15.51798  8.415264 0.8673273
## Final Assembly-Top Line   -3.8719500 -12.91787  5.173966 0.6489765
## Machining-Top Line        -4.0217316 -15.98835  7.944890 0.8096473
## Final Assembly-Finishing  -0.3205923 -12.28721 11.646030 0.9999850
## Machining-Finishing       -0.4703740 -14.77322 13.832475 0.9999660
## Machining-Final Assembly  -0.1497817 -12.11640 11.816840 0.9999993
```

Figure B36: Tukey multiple pairwise comparison for formaldehyde. Groups for this test were all the "in factory" production lines described in section 2.1 of this appendix.

The residuals plot in Figure B37 was used to visually identify variance patterns that would be a sign of a violation of the homoscedasticity assumption (e.g., funnel or cornucopia shapes) required for the ANOVA test. No evidence of a violation of the assumption was found for the formaldehyde exposure data.

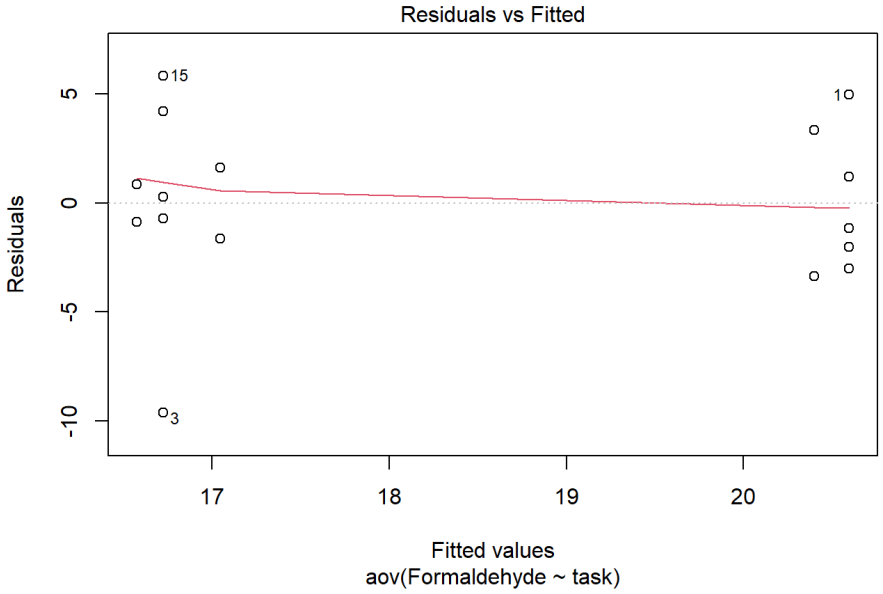

Figure B37: Residuals plot for formaldehyde.

Levene's test is a probabilistic approach for detecting violations of the homoscedasticity violation. This test was used to confirm the visual assessment described in the previous subsection (i.e., the residuals plot approach). The null hypothesis of Levene's test is that the variances are homogeneous. The *p*-value of 0.6326 shown in Figure B38 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the homogeneity of variance assumption for the formaldehyde exposure data.

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  0.6596 0.6326
##       11
```

Figure B38: Output of Levene's test for formaldehyde.

The Q-Q plot in Figure B39 was used to visually assess the normality of the formaldehyde exposure data. No major deviations from the line of normality (i.e., the 1:1 line) were observed. No violations of the normality assumption were found.

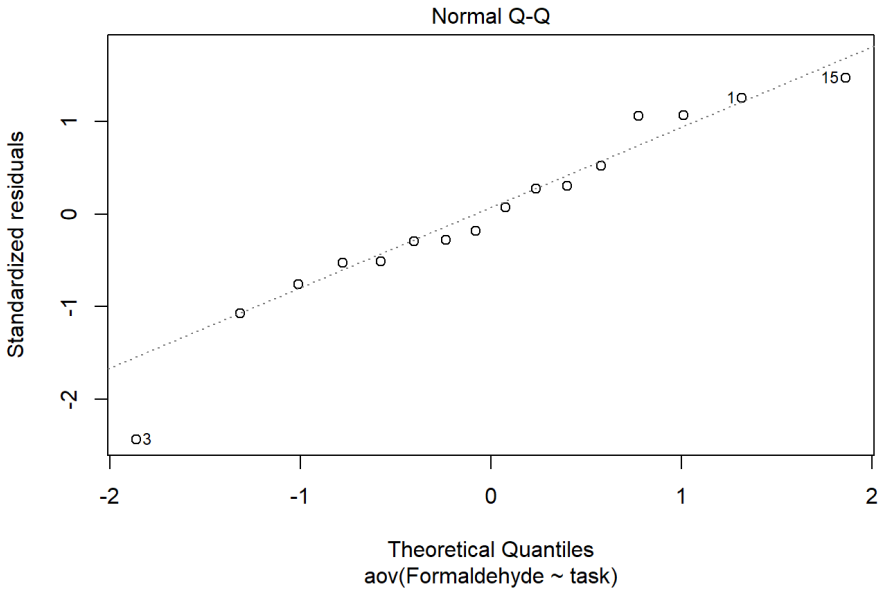

Figure B39: Q-Q plot for formaldehyde.

The Shapiro-Wilk test is a probabilistic approach for detecting violations of the normality assumption required for the ANOVA test. This test was used to confirm the visual assessment described in the previous subsection (i.e., the Q-Q plot approach). The null hypothesis of the Shapiro-Wilk test is that data comes from a normal distribution. The *p*-value of 0.4047 shown in Figure B40 means that we did not have enough evidence to reject the null hypothesis at a significance level of 0.05. In other words, we did not find a violation of the normality assumption for the formaldehyde exposure data.

```
##
##  Shapiro-Wilk normality test
##
## data:  aov_residuals
## W = 0.94428, p-value = 0.4047
```

Figure B40: Output of Shapiro-Wilk's normality test for formaldehyde.

9. Noise monitoring implementation

9.1. Data acquisition and processing

The microphone on the AirPen produces an analog signal with voltages that oscillate around zero volts (i.e., voltages can be positive and negative). A DC offset of ½ of the reference voltage (V_{ref} = 3.3V) is added to make all voltages positive (i.e., the conditioned signal oscillates around +1.65V). The signal is then amplified (8× gain) and sampled by the AirPen's internal analog-to-digital converter (ADC) at 50000 Hz. The ADC converts the analog signal to 14-bit resolution digital values, which are logged to the SD card in chunks of 4096 samples (i.e., columns) per observation (i.e., per row). This sampling configuration gives a sampling window of 0.082 seconds per observation.

The raw time-domain noise data recorded by the AirPen were transformed to loudness using an R script on a personal computer. First, the DC offset of each unit was subtracted to recover the original shape of the microphone signal with positive and negative values. Then, the values were normalized by dividing them by the ADC full scale. For each observation, the vector of 4096 normalized values was transformed to a 2048-bin power spectrum in the 0 – 25 kHz frequency range using Fast Fourier Transform (FFT). The power spectrum (i.e., frequency domain data) was A-weighted by multiplying each bin by the A-weighting coefficient corresponding to its frequency.[87] The uncalibrated A-weighted loudness was calculated Equation B2:

$$Loudness\ in\ dBA = 10\log_{10}\left(\frac{2}{N\Delta t}\sum_{k=0}^{N/2}|X_A[k]|^2\right) \qquad (B2)$$

Where $X_A[k]$ was the k-th bin of the A-weighted power spectrum, N was the number of samples per observation (i.e., 4096), and Δt was the sampling window (i.e., 0.082 seconds).

9.2. Calibration

Individual calibration curves were developed for each AirPen. A Larson Davies Spark 703+ dosimeter (Depew, NY, USA) was used as the reference instrument for calibration. Each AirPen was placed within 5 cm of the microphone of the dosimeter. Two speakers (532-LR, Aperion Audio, Wilsonville, OR, USA) were placed at a distance of 30 cm from the noise measuring instruments, aiming directly at the microphones. An amplifier (TX-SR605, Onkyo, Osaka, Japan) connected to a laptop was used to generate a 1 kHz tone. Calibration was carried out in the 50 – 110 dBA range at 9 evenly spaced discrete loudness levels. Each loudness level was sampled at least twice. A second-degree polynomial was fitted to the calibration data using the ordinary least squares method. A silent sample was also recorded to measure the deviation of the true DC offset

of each AirPen from the theoretical DC offset (i.e., ½ of V_{ref}). Individual calibration coefficients and true DC offsets were saved and used to calculate calibrated loudness levels of noise measured during the facility-wide personal sampling campaign.

9.3. Validation

We assessed the noise monitoring performance of the AirPens by collocating them with calibrated Larson Davis Spark 703+ dosimeters (Depew, NY, USA) in three tests that were completed on two different days in a variety of environments within the Colorado State University Powerhouse Energy Campus. Spark 703+ dosimeters meet the IEC 61252 standard for personal exposure noise dosimeters, applicable to instruments for measurement of A-weighted steady, intermittent, fluctuating, irregular, or impulsive sound.[87] This type of instrument is considered a gold-standard for industrial hygiene. A researcher wore one AirPen and one Spark 703+ dosimeter on the same arm, with the microphones positioned within 10 centimeters of each other. For the first test, one set of samplers (i.e., one AirPen and one dosimeter) was worn on the right arm. For the second test, one set of samplers was worn on the left arm. For the third test, AirPens and Spark 703+ dosimeters were worn on both arms (i.e., one set per arm, 4 samplers in total). All tests included a mixture of different noise sources. Normal office space background noise and conversational noise were measured in areas where students and researchers work at cubicles. Large engine (e.g. fuel cell generators, gas turbines) tests were recorded to get quasi-steady loud noise. Irregular and fluctuating noise was measured at a machining and fabrication area.

AirPens measured noise in 30-second intervals, and dosimeters were set to 1-Hz data logging. Measurements from colocated AirPens and dosimeters were merged by date/time based on their internal clocks, creating a set of paired measurements at 30-second resolution. Clocks were set

on both instruments before each test. The AirPen used GPS to adjust its clock, and the dosimeter clock was set by an NTP-connected personal computer.

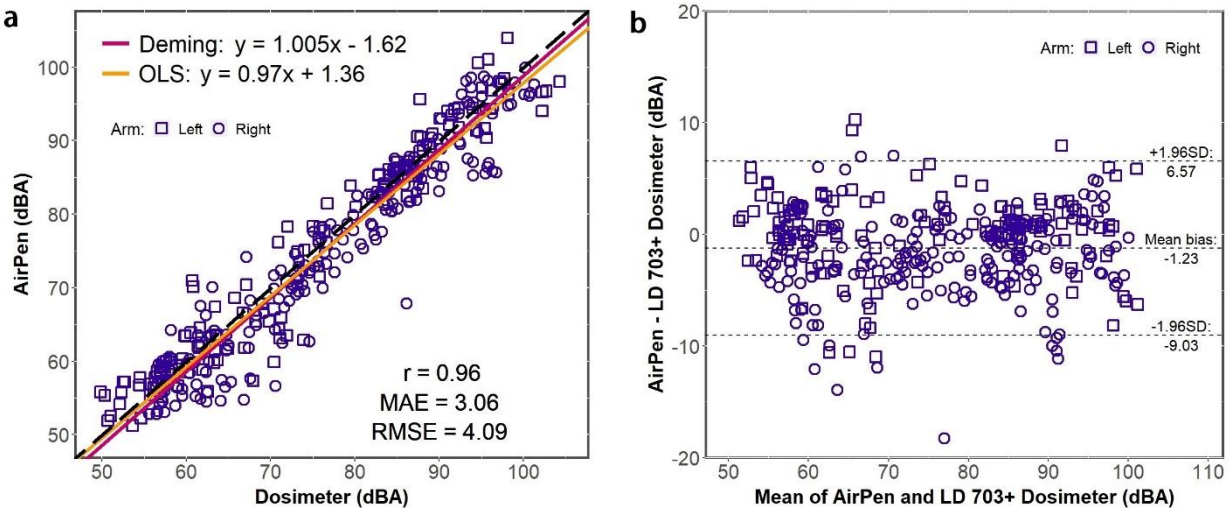

Figure B41: AirPen noise monitoring validation results. **a)** Regression plot of AirPens versus Larson Davis Spark 703+ dosimeters at 30-second resolution and error metrics. Dashed black line is the identity (1:1) line and colored lines are regression fits. **b)** Bland-Altman plot shows the error distribution as a function of noise levels. Points within the upper and lower dashed lines are the central 95% of the error distribution.

Figure B41 shows regression fits and error metrics of AirPen-measured A-weighted loudness, relative to Spark 703+ dosimeters, as well as a Bland-Altman plot to portray the distribution of errors as a function of loudness. We present a Deming regression as well as the usual ordinary least squares (OLS) regression because the Deming method is more appropriate when both axes (i.e., instruments) are subject to measurement errors. The 30-second resolution AirPen measurements were highly correlated with the corresponding Spark 703+ dosimeter measurements (Pearson's $r = 0.96$), and negatively biased by 1.23 dBA, on average. Mean absolute error (MAE) and root mean squared error (RMSE) were 3.06 dBA and 4.09 dBA, respectively. Table B2 shows the time-weighted averages (TWA) of A-weighted noise measured by AirPens and Spark 703+ dosimeters for three different tests. Bias of AirPen-measured TWA noise relative to Spark 703+ dosimeters ranged from -2.6 to +0.7 dBA.

Table B2: Time-weighted averages of A-weighted noise level measured by AirPens and Larson Davis Spark 703+ dosimeters, as well as biases for each test.

	TWA Noise [dBA]					
	Right Arm			*Left Arm*		
Test	*AirPen*	*Dosimeter*	*Bias*	*AirPen*	*Dosimeter*	*Bias*
Test 1 – right arm only (38 minutes)	76.2	78.8	-2.6	N/A	N/A	N/A
Test 2 – left arm only (14 minutes)	N/A	N/A	N/A	57.5	56.8	+0.7
Test 3 – both arms (57 minutes)	75.6	77.7	-2.1	78.3	78.2	+0.1

APPENDIX C

1. Community organizations and super neighborhoods

The community organizations that participated in this study and their corresponding super neighborhoods are listed in Table C1.

Table C1: Organizations that hosted samplers and their corresponding super neighborhoods or areas.

Organization	*Super neighborhood / Area*
Sunnyside Community Redevelopment Organization (SCRO)	Sunnyside
Achieving Community Tasks Successfully (ACTS)	Pleasantville
Coalition of Community Organizations (COCO)	Fifth Ward
NorthEast Houston Redevelopment Council (NEHRC)	Trinity Gardens
Environmental Community Advocates of Galena Park (ECAGP)	Galena Park
Community In- Power and Development Association Incorporate (CIDA)	Port Arthur
Texas Southern University Bullard Center for Climate & Environmental Justice	Third Ward
Environmental Defense Fund	Greater Heights

2. Sample analysis

2.1. Analytical method

The contents of each tube were extracted using 3 mL of a mixture of 70% acetonitrile and 30% deionized water drawn through the cartridge using a vacuum manifold. The extract was collected in a 5 mL volumetric flask. Once extraction was complete, the volumetric flask was filled to a final extraction volume with the 70% acetonitrile:30% deionized water. Each extract was allowed to equilibrate for 20 min following the instrument manufacturer's instructions and then analyzed immediately on an Agilent 1260 Infinity HPLC (Santa Clara, CA, USA) equipped with

a dual channel pump and UV detector set to monitor light intensity at a wavelength of 360 nm. The separation was conducted on a Waters Nova-Pak (Milford, Massachusetts, USA) C-18 4 μm column (3.9 x 150 mm). The eluents were 60% deionized water:30% acetonitrile:10% tetrahydrofuran (A) and 40% water:60% acetonitrile (B). The complete run time was 32 minutes and included 2 steps. For the first 25 minutes, a linear gradient from 100% A to 100% B was conducted. Then, for the final 5 minutes, a re-equilibration step was done to return to the starting conditions of 100% A. The flowrate during each step was 1.5 mL min^{-1}. An injection volume of 10 μL was used for each sample.

2.2. Limit of detection

$$LOD = \bar{x}_{blank} + 3s_{blank} \qquad (C1)$$

The formaldehyde limit of detection (LOD) was calculated using Equation C1 where $\bar{x}_{blank}$ was the mean and s_{blank} was the standard deviation of the peak areas for the blank BPE-DNPH cartridges collected during the field campaign. Blank cartridges were briefly installed in samplers, removed, capped, and sent to the laboratory for analysis along with sample cartridges. No air was sampled through blank cartridges. The peak area LOD can be converted to mass units (in ng) using the calibration curve for formaldehyde. The formaldehyde LOD was 36.6 ng (calculated from 8 blanks). The volumetric flow rate through the gas cartridge is not actively controlled on the AirPen, so we provide LOD equivalent concentrations for some relevant measured volumes. The formaldehyde LOD is equivalent to concentrations of 1.88, 0.39, and 0.22 μg m^{-3} for the minimum, median, and maximum measured volumes, respectively. In parts per billion (ppb) these equivalent LODs are 1.53, 0.32, and 0.18 at 1 atm and 25°C (study location was at sea level and weather was warm).

3. Flow control analysis

The first and last 5 minutes of each sample were disregarded to account for the initial stabilization and the effect of low battery state-of-charge on pump voltage at the end of some samples. Individual observations with flow rate through the inlet deviating from the target (250 mL min^{-1}) by more than 10% were flagged as point anomalies. Flow rate through the BPE-DNPH is passively controlled by an orifice in the manifold, so a target is not available, and manufacturing variability is expected. Therefore, we only flagged flow rates through the BPE-DNPH cartridge as potential blockages when they were lower than the median (24 mL min^{-1}) by at least two median absolute deviations (9 mL min^{-1}). Flow rates through the BPE-DNPH cartridge were flagged as potential leaks when they were higher than the median by at least two median absolute deviations. The median absolute deviation was used because it is more robust to outliers. For pump voltage, a 10-minute rolling mean (still at 30-resolution) was used to catch only longer-term trends. Changes of the pump voltage rolling mean larger than 2.5 V h^{-1} were flagged as point anomalies. Samples with 15 or more minutes of anomalous data of any kind (i.e., ≥ 30 point anomalies) were flagged as potentially anomalous samples and assessed graphically to confirm the anomaly.

3.1. Example of adequate flow control (non-anomalous reference)

The example below (Figure C1) corresponds to one of the samples collected in the field and represents a good reference of how AirPens performs when everything works as intended. The anomaly detection algorithm did not flag any of the timestamps as anomalous. Upon visual inspection, we found the pump voltage, flow rate through the inlet, and flow rate through the BPE-DNPH cartridge to be steady throughout the duration of the sample, apart from some very short deviations. The flow rate through the inlet was practically equal to the target most of the

time. The flow rate through the BPE-DNPH cartridge was slightly higher than the median at ~28 mL min^{-1}.

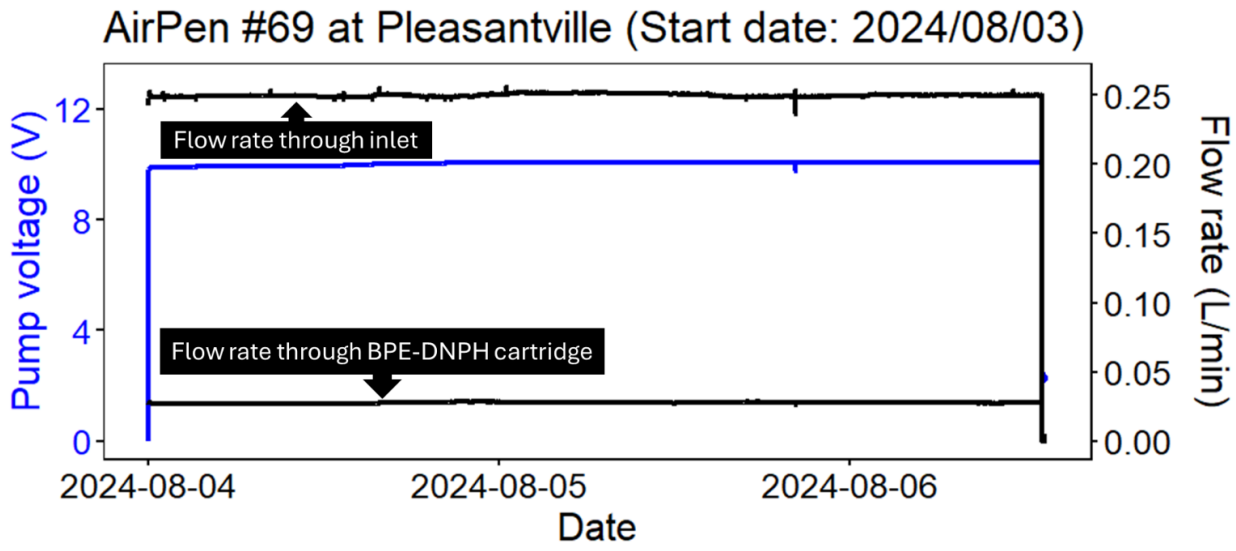

Figure C1: Example of a sample with flawless flow control. Blue line shows the pump voltage (V). The top black line shows the flow rate through the inlet (L/min) and the bottom black line shows the flow rate through the BPE-DNPH cartridge (L/min).

3.2. Samples flagged as anomalous (blockages)

The following samples (Figures C2 and C3) were flagged by the flow analysis algorithms, and, upon visual inspection, we found that the flow rate through the BPE-DNPH cartridge was nearly zero throughout the sampling period. Considering that pump voltage and flow rate through the inlet were normal, these data are consistent with the existence of a blockage in the BPE-DNPH cartridge flow path.

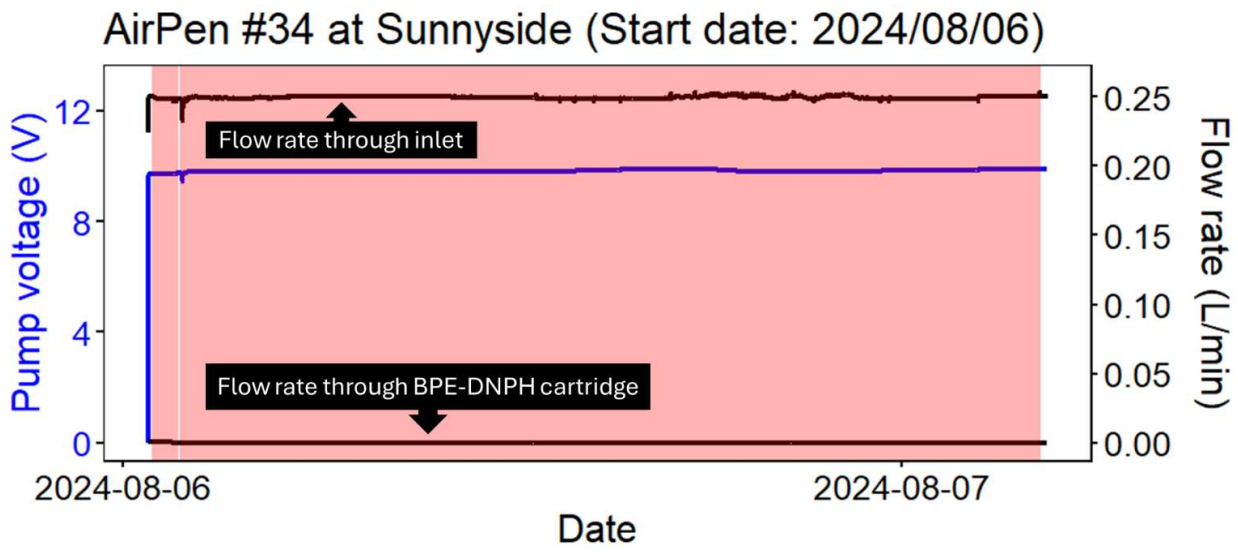

Figure C2: Invalid sample with a blockage in the flow path. Blue line shows the pump voltage (V). The top black line shows the flow rate through the inlet (L/min) and the bottom black line shows the flow rate through the BPE-DNPH cartridge (L/min). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm.

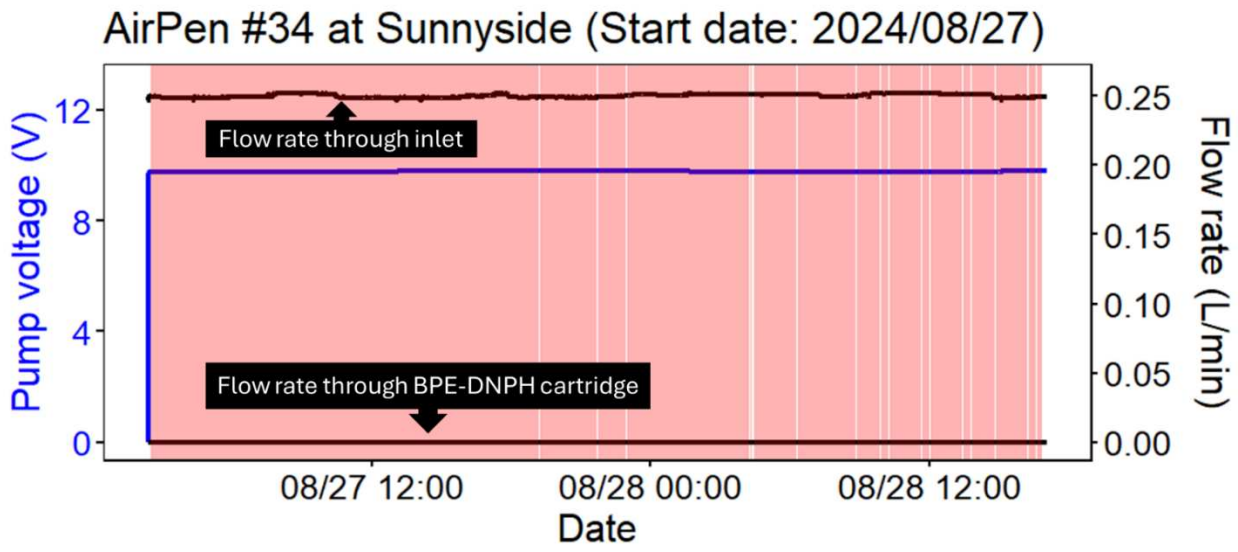

Figure C3: Invalid sample with a blockage in the flow path. Blue line shows the pump voltage (V). The top black line shows the flow rate through the inlet (L/min) and the bottom black line shows the flow rate through the BPE-DNPH cartridge (L/min). Areas highlighted in red correspond to timestamps flagged as anomalous by the algorithm