

# W-Air: Enabling Personal Air Pollution Monitoring on Wearables

BALZ MAAG, ETH Zurich, Switzerland

ZIMU ZHOU, ETH Zurich, Switzerland

LOTHAR THIELE, ETH Zurich, Switzerland

Accurate, portable and personal air pollution sensing devices enable quantification of individual exposure to air pollution, personalized health advice and assistance applications. Wearables are promising (e.g., on wristbands, attached to belts or backpacks) to integrate commercial off-the-shelf gas sensors for personal air pollution sensing. Yet previous research lacks comprehensive investigations on the accuracies of air pollution sensing on wearables. In response, we proposed W-Air, an accurate personal multi-pollutant monitoring platform for wearables. We discovered that human emissions introduce non-linear interference when low-cost gas sensors are integrated into wearables, which is overlooked in existing studies. W-Air adopts a sensor-fusion calibration scheme to recover high-fidelity ambient pollutant concentrations from the human interference. It also leverages a neural network with shared hidden layers to boost calibration parameter training with fewer measurements and utilizes semi-supervised regression for calibration parameter updating with little user intervention. We prototyped W-Air on a wristband with low-cost gas sensors. Evaluations demonstrated that W-Air reports accurate measurements both with and without human interference and is able to automatically learn and adapt to new environments.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**;

Additional Key Words and Phrases: Wearables, Air Pollution, Sensor Array, Calibration

## ACM Reference Format:

Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. W-Air: Enabling Personal Air Pollution Monitoring on Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 24 (March 2018), 25 pages. <https://doi.org/10.1145/3191756>

## 1 INTRODUCTION

**Air pollution monitoring.** Air pollution affects human health, productivity and comfort. Atmospheric pollutants outdoors such as ozone (O<sub>3</sub>) contribute to respiratory symptoms when people engage in outdoor exercises and activities [23]. Continuous ozone monitoring is not only important due to ozone's toxicity but also due to its strong impact on various other major pollutants such as nitrogen oxides (NO<sub>x</sub>) [8]. According to the World Health Organization (WHO), 11.6% of the global deaths in 2012 were linked to air pollution [55]. Providing air pollution information to individuals enables them to understand and improve the air quality of their living environments.

There is a growing demand to increase the spatio-temporal resolution of air pollution monitoring. Governmental institutions deploy expensive high-end air pollution sensors in a few stations across a city. The measurements only suffice to estimate the average pollution exposure experienced by the majority of the population for urban planning and policy making. In recent years low-cost sensors have been deployed by researchers and agencies in

---

Corresponding Author: Zimu Zhou.

Authors' addresses: Balz Maag, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, Switzerland, balz.maag@tik.ee.ethz.ch; Zimu Zhou, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, Switzerland, zimu.zhou@tik.ee.ethz.ch; Lothar Thiele, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, Switzerland, thiele@tik.ee.ethz.ch.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Association for Computing Machinery.

2474-9567/2018/3-ART24 \$15.00

<https://doi.org/10.1145/3191756>

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 2, No. 1, Article 24. Publication date: March 2018.

static stations [56] or on mobile vehicles [15] to build air pollution maps for citizens. However, these maps have in general low accuracy and low spatio-temporal resolution and, therefore, can be misleading when assessing personal exposure for quantitative studies and applications. Due to the complex heterogeneity of air pollutants [27] and the diverse moving patterns of individuals [18], *personal* air pollution sampling is necessary for meaningful personal exposure analysis [33, 36].

**Motivation.** Personal air pollution monitoring for quantitative health and well-being applications requires accurate, convenient, quasi-continuous collection of heterogeneous data. For instance, researchers record both individual micro-environment (*e.g.*, temperature, wind speed, noise, air pollution) and psychological states (*e.g.*, skin temperature, heart rate) via a set of wearable sensors to investigate the impact of urban environments on citizen’s health and quality of life [28, 29, 41]. Simultaneous sensing of biological and environmental data enables personalized advice and assistance applications that promote a healthier lifestyle and improved health-care prevention.

To design air pollution sensing devices for the above studies and applications, we primarily focus on integrating low-cost gas sensors into wrist-worn devices. (i) Many users wear wristbands or smart-watches most of the day, providing long-term closeness to the body and exposure to the ambient air. (ii) Biological parameters are commonly measured by wearables such as wristbands [29, 41]. Integrating environmental sensors into wearables can improve the compactness and usability of the infrastructure for environment-related physiological applications.

**Challenges.** Despite various portable air pollution sensing devices [3, 5, 18, 34, 60], there is a void in air pollution sensing on wearables. Due to their small size, metal oxide (MOX) sensors have been widely adopted to measure a wide spectrum of important air pollutants [21] on portable devices [32, 36]. However, similar to other low-cost gas sensors, MOX sensors suffer from low selectivity, *i.e.*, they are *cross-sensitive* to various substances in the air [25]. Since human beings can emit multiple gases through natural skin oils [54], cosmetics [53], textiles [38] or respiration [7], these gases may interfere with the measurements of the *ambient* atmospheric pollutants when the sensors are placed close to the human body (*e.g.*, on wristbands, attached to belts or backpacks). Hence it is crucial to investigate and filter the *human interference* to acquire accurate measurements of ambient air pollution.

In this paper, we design and implement *W-Air*, the *first-of-its-kind personal air pollutant monitoring platform for wearables with low-cost gas sensors*. As a proof-of-concept, we focus on monitoring concentrations of ambient O<sub>3</sub> and CO<sub>2</sub>, two important outdoor and indoor gas pollutants. We aim to answer the following questions.

- *Does the wearable setting introduce new challenges for accurate air pollution sensing?* Through measurement studies, we observe that the measurements of low-cost MOX gas sensors significantly deviate from the ground truth when there is close human presence and the deviation is *non-linear*. A linear calibration model, which is typically provided by the sensor manufacturer, yields measurement errors that make it impossible to draw any conclusions about the actual O<sub>3</sub> and CO<sub>2</sub> concentration when there is human interference. We also note that such human interference is not restricted to the wrist-worn setting. Attaching the gas sensors to other popular wearable settings, such as belts or backpacks, still yield notable interference from human emissions.
- *How to enable accurate air pollution monitoring on wearables?* A key finding is that the human interference can be characterized by increasing VOC concentrations. Different MOX gas sensors react differently to these VOC concentrations due to their individual specific cross-sensitivities. Therefore we design a neural network based calibration framework by jointly considering measurements from two different low-cost MOX sensors to eliminate the non-linear interference of VOC emissions caused by human beings. Since *W-Air* is expected to measure ambient O<sub>3</sub> and CO<sub>2</sub> concentration accurately both with and without human interference, large amounts of measurements in both cases are needed for training the calibration parameters. To improve the usability of *W-Air*, we apply a neural network architecture with shared layers

and a semi-supervised regression technique to boost the training process with *fewer samples* and *little user intervention* compared to traditional neural network architectures.

**Contributions and road-map.** The main contributions of this work are summarized as follows.

- We discover and characterize the human interference on the measurements of low-cost MOX sensors, a crucial yet largely overlooked issue to obtain accurate ambient gas measurements on wearables (*e.g.*, wrist-worn, attached to a belt or a backpack).
- We propose an effective sensor-fusion calibration scheme to recover ambient outdoor O<sub>3</sub> and indoor CO<sub>2</sub> concentrations from low-cost gas sensor readings during human interference situations. Our calibration method is based on a neural network that uses measurements from two different MOX gas sensors and a temperature sensor to accurately estimate O<sub>3</sub> and CO<sub>2</sub> concentrations. We further utilize a shared layer architecture to bootstrap the supervised training process of our neural network with fewer samples compared to typical neural network architectures. Additionally, we apply semi-supervised regression for parameter updating with little human intervention. W-Air sets a new standard for portable air pollution monitoring with easy maintenance.
- We prototype the above design using COTS MOX gas sensors integrated on a wristband platform. To the best of our knowledge, W-Air is the first working air pollutant monitoring platform using COTS gas sensors for wristbands and smart-watches. Evaluations show that W-Air is able to measure ambient O<sub>3</sub> and CO<sub>2</sub> concentrations with an error of around 4.3 ppb and 64 ppb respectively. Overall, W-Air achieves a data quality that is sufficient for personal air pollution monitoring [47, 49].

In the rest of the paper, we first review relevant literature (Sec. 2), present the measurement study (Sec. 3) and the exploratory calibration (Sec. 4) on the human interference. Then we elaborate on the design (Sec. 5) and evaluation (Sec. 6) of W-Air. We discuss limitations and future work in Sec. 7 and finally conclude this work in Sec. 8.

## 2 RELATED WORK

W-Air is proposed to meet the need for applications in *crowdsourced environment sensing* and *personal environment sensing*. The design of W-Air is built upon previous research on *portable air pollution sensing devices* and *gas sensor calibration*. We review the closely relevant works as follows.

### 2.1 Crowdsourced Environment Sensing

Mobile crowdsourcing, or participatory sensing, has been widely adopted for environment sensing. In crowdsourced environment sensing, unprofessional users take measurements of the environment “in the form of an open call” [17] with their smartphones or other portable devices to cover a large spatio-temporal range. For example, Ear-Phone [39] is an end-to-end crowdsourced noise sensing and mapping system that builds an urban noise map by measuring noise from smartphones. It leverages compressive sensing to construct accurate noise maps from the sparse and random crowdsourced noise measurements. Overeem *et al.* [35] design techniques to infer ambient temperature from crowdsourced smartphone battery temperature and air temperature reported by meteorological stations. Combined with location and time information, they achieve an average estimation accuracy of 1.45°C. Atmos [30, 31] is a crowdsourced weather data application that not only automatically samples smartphone sensors (GPS, temperature, light, pressure), but also allows manual input for current and future weather condition estimation. The Atmos application shows an average accuracy of less than 2.7°C for ambient temperature estimation using such a hybrid (automatic sensing and manual user input) crowdsourcing approach.

Crowdsourced air pollution sensing is also of growing research interest because static air pollution monitoring stations are sparsely deployed in cities and limited in spatial resolution [15]. Compared with other environmental

Table 1. Comparison of portable air pollution sensing devices. W-Air is designed for wearable settings, evaluated in both indoor and outdoor scenarios, and targets both crowdsourced and personal environment sensing applications.

System	Pollutant	Technology	Usage	Scenario	Application
Budde <i>et al.</i> [3]	Dust (PM <sub>2.5</sub> , PM <sub>10</sub> )	Optical	Carriable	O	C
AirSense [60]	Dust (PM <sub>2.5</sub> )	Optical	Carriable	I + O	C + P
MyPart [48]	Dust (PM <sub>10</sub> )	Optical	Wearable	I + O	C + P
MAQS [18]	Gas (CO <sub>2</sub> )	Optical	Carriable	I	P
CitiSense [32]	Gas (CO, NO <sub>2</sub> , O <sub>3</sub> )	MOX	Carriable	I + O	C
Common Sense [5]	Gas (CO, NO <sub>x</sub> , O <sub>3</sub> )	N/A	Handheld	O	C
Oletic <i>et al.</i> [34]	Gas (CO, NO <sub>2</sub> , SO <sub>2</sub> )	Electrochemical	Handheld	O	C
Piedrahita <i>et al.</i> [36]	Gas (CO, CO <sub>2</sub> , NO <sub>2</sub> , O <sub>3</sub> )	MOX, Optical	Carriable	I + O	C + P
<b>W-Air</b>	<b>Gas (CO<sub>2</sub>, O<sub>3</sub>)</b>	<b>MOX</b>	<b>Wearable</b>	<b>I + O</b>	<b>C + P</b>

Note 1: I - indoor; O - outdoor; C - crowdsourced air pollution sensing; P - personal exposure monitoring.

Note 2: The usage of the device is partially determined by its compactness. Wrist-worn devices are smaller in size than carriable and handheld designs.

data, air pollution (e.g., gases and dust) can hardly be measured by smartphone sensors. Hence the first step for crowdsourced air pollution sensing is to design portable and accurate sensing devices suitable for crowdsourced users, which partially motivates our work.

## 2.2 Personal Environment Sensing

Due to the complex spatial heterogeneity of pollutants [27], personal environment sensing is important for quantitative studies such as personal exposure assessment. For example, Oglesby *et al.* [33] report that personal air pollution sampling is necessary for short-term analysis of personal ultrafine particle exposure. In many personal environment sensing applications, the sensing device needs to measure both biological responses and environmental data. Nakayoshi *et al.* [28] investigate outdoor thermal physiology by deploying wearable sensors to record individual microclimate (temperature, humidity, wind speed and radiation) and psychological states (skin temperature, heart rate). Project ESUM [41] studies the impact of urban morphology on citizen's social potential (e.g., perception). Environmental conditions (e.g., noise, temperature, illumination, air pollution) and skin conductance responses are recorded when participants walk around the city. Project CONVERGENCE [29] integrates low-power environmental sensors and biological sensors on wearables for new generations of human-machine interfaces and health-care and lifestyle applications.

In these personal environment sensing studies and applications, participants have to carry a set of sensors to collect environmental and biological data. The complexity and bulkiness of the sensing infrastructure may discourage user engagement and even induce bias for physiological studies. Note that biological or physiological parameters are commonly measured by wearables such as wristbands [41]. Therefore our work explores environment sensing on wearables to improve the compactness and usability of the sensing devices for environment-related physiological studies and personalized health advice and assistance applications.

## 2.3 Portable Air Pollution Sensing Devices

The need for crowdsourced and personal air pollution monitoring has fostered various portable air pollution sensing devices [3, 5, 18, 32, 34, 36, 48, 60]. Table 1 summarizes the target pollutants, sensor technologies, device usage, scenarios and applications of existing portable air pollution sensing devices. Mainstream COTS air pollution sensor technologies include metal oxide (MOX), electrochemical or optical approaches, where MOX sensors are

the most popular for their small sizes [21]. MOX sensors are primarily used for monitoring gaseous pollutants such as CO, NO<sub>2</sub>, O<sub>3</sub> and VOCs [5, 32, 36], while optical sensors are preferable for dust monitoring (*e.g.*, PM<sub>2.5</sub>, where PM stands for particulate matter, and PM<sub>2.5</sub> means fine particles with a diameter of 2.5 micrometers or less) [3, 60]. Sensing devices with MOX sensors vary in size depending on mechanical designs and functionalities, while platforms integrated with COTS electrochemical or optical sensors can only be handheld or attached to backpacks limited by the form factor of the sensors.

Some works focus on platform integration and qualitative validation. Common Sense [5] and CitiSense [32] design sensor nodes with wireless connection and interfaces for crowdsourced gas pollutant monitoring. AirSense [60] integrates dust sensors for personal PM<sub>2.5</sub> monitoring and conduct feasibility studies in various contexts. Other works investigate the reliability and accuracy of sensors and emphasize more on quantitative evaluation. Oletic *et al.* [34] conduct outdoor field tests for gas pollutant sensing. MyPart [48] is a wrist-worn PM<sub>10</sub> sensing device that works both indoors and outdoors with validated accuracy.

The closest to our work is [36], where the authors measure a set of gas pollutants (CO, CO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>) using low-cost MOX sensors, calibrate and validate the sensor readings in both indoor and outdoor scenarios. W-Air is also an accurate, multi-gas sensing device that works both indoors and outdoors. However, W-Air primarily focuses on a wrist-worn setting, and identifies the human interference problem. We demonstrate that the human interference problem is generic for other sensor placement (*e.g.*, attached to a belt or backpack), and propose effective calibration schemes to filter such interference. W-Air can also deal with insufficient and imbalanced data, which is important for sensor deployments in the wild, yet largely overlooked in previous studies.

## 2.4 Calibration of Gas Sensors

Although manufacturers had calibrated their products in labs, field testing and calibration are important for low-cost gas sensors to yield meaningful measurements. Low-cost gas sensors also drift over time and need periodic re-calibration.

Univariate linear regression techniques such as ordinary least squares [14] are commonly applied if the sensor is only sensitive to the target pollutant. If the gas sensor is also sensitive to other air pollutants, *i.e.*, cross-sensitive, the sensor is usually accompanied by other sensors to form a sensor array [25]. Instead of calibrating each cross-sensitive sensor individually, researchers propose to jointly calibrate a sensor array via linear regression techniques [20, 25] or artificial neural networks [4, 45].

Previous gas sensor calibration research assumes the calibration is performed by professionals. For sensor platforms designed for unprofessional end users, it is crucial to involve minimal efforts from end users for calibration parameter training and updating. W-Air advances existing sensor array calibration research by (i) accounting for the non-linear interference from close human presence and (ii) leveraging multi-task and semi-supervised learning frameworks for sensor array calibration with limited and imbalanced training data.

## 3 MEASUREMENT STUDY

In this section, we show through field studies the human interference on the measurements of COTS MOX sensors. Such human interference imposes enormous errors when using compact gas sensors to monitor ambient atmospheric pollutants. We mainly focus our measurements study on O<sub>3</sub> for outdoor environments and CO<sub>2</sub> for indoor environments due to (i) their importance for air quality assessments and (ii) the availability of promising low-cost sensor technology.

### 3.1 Measurements

Despite laboratory and in-field calibration from manufacturers, it is still indispensable to conduct pre-deployment testing and calibration of low-cost gas sensors in the target environments [25]. Many existing portable air

pollution monitoring devices aim to measure coarse-grained air quality indices, and their measurements are not validated by highly accurate gas sensors [5, 18, 32]. Other works either perform data validation in chambers [3] or directly adopt calibration parameters from manufacturers [34]. MyPart [48] conducts in-field validation for *customized* dust sensors. In [36], the authors report low correlation of O<sub>3</sub> measurements between the metal oxide (MOX) gas sensors and the ground-truth, yet without in-depth investigations.

**3.1.1 Measurement Settings.** We deploy different sensors, depicted in Fig. 1, in different indoor and outdoor settings.

**Locations.** Outdoors we deploy the sensors on the roof of our academic building in the city-center and on a balcony in a residential area. Indoors the sensors are deployed either in an office environment, a living-room or a bed-room.

**Low-cost sensors.** The sensors are two low-cost MOX sensors, namely a *MICS-OZ-47* O<sub>3</sub> sensor [43] and a *CCS811* VOC sensor [2]. The two sensors are connected to the *Thunderboard Sense* [22], a compact multi-sensor development platform, that acts as a wearable device. Note that the *MICS-OZ-47* O<sub>3</sub> sensor is only providing raw and uncalibrated analog-digital-converter (ADC) values and the *CCS811* sensor already calibrated VOC concentration values within [0,1187] ppb (parts-per-billion = 10<sup>-7</sup>%). Additionally we use the *Si7021* temperature and relative humidity sensor to monitor environmental conditions.

**Ground-truth.** As ground-truth reference devices we use a *SM-50* O<sub>3</sub> measurement unit [1] in outdoors environments and a *Telaire 6713* CO<sub>2</sub> measurement unit [42] in indoor environments. The *SM-50* sensor provides highly accurate ozone measurements within [0,150] ppb, *i.e.*, typical ambient outdoor concentrations. The *Telaire 6713* measures typical indoor CO<sub>2</sub> concentrations within [400,5000] ppb with high accuracy. Although it is conceivable to also integrate the *Telaire 6713* into wearables due to its form factor, its power consumption of 125 mW is almost 10 times higher than the one of a MOX sensor.

**Approach.** We use the VOC readings to approximate the CO<sub>2</sub> concentration because the *CCS811* VOC sensor is not directly measuring CO<sub>2</sub>. This is a widely used approach of available low-cost CO<sub>2</sub> sensors due to the high correlation between VOC and CO<sub>2</sub> [16]. Furthermore, CO<sub>2</sub> is mainly used to assess indoor air quality [19]. The sensors on the wristband are sampled at 2 Hz, the O<sub>3</sub> reference sensor every 1 min and the CO<sub>2</sub> every 5 sec. Additionally we smooth the wristband sensor readings over a sliding window of 3 sec. We power all the sensors and collect their measurements using a laptop to ensure long-term measurements.

We use the measurements from the MOX sensors to reveal a major problem of state-of-the-art MOX gas sensors: human beings act as a VOC source due to emissions from the skin, clothes or respiration [53]. These VOC emissions can be detected by state-of-the-art MOX sensors and, therefore, interfere with the assessment of the ambient air quality. Related work has shown that the emissions are detectable up to 1 m away of a human being [12] and, thus, potentially affect typical wearable air quality monitors. Therefore, during the measurements with human interference a user is wearing the device while staying between 1.5 m and 2 m away from the reference sensors, depicted as  $S_H$  in Fig. 1. We consider three popular ways of utilisation of wearable air quality monitors, namely (i) a wrist-band [48], (ii) attached to a belt [6] and (iii) attached to a backpack [3, 18], as shown in Fig. 2. Further, we place a second wearable device ( $S_N$ ) next to the reference sensor to highlight that the human interference problem is in particular severe when the sensors are close to a human being.

**3.1.2 Data Collection.** We collected data consisting of approximately 100 hours each in indoor and outdoor environments distributed over 21 days between April and October, 2017. The measurements were conducted at different times during the day, *e.g.*, morning, midday, afternoon, evening and night, and in various weather conditions, *e.g.*, heavy and light rain, cloudy, sunny and windy. We collected both measurements with and without human presence. During human presence situations a user is wearing the wearable device in one of the three different usages shown in Fig. 2 and sitting or standing next to the corresponding reference device as indicated in

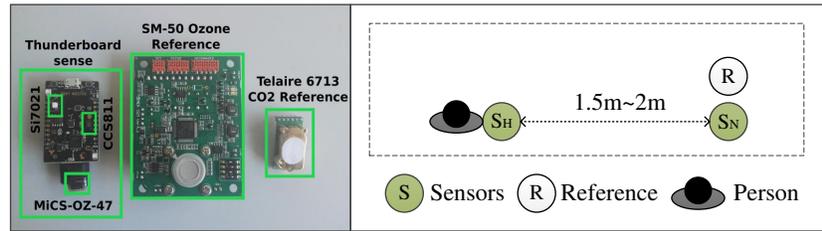


Fig. 1. Settings of measurement studies: sensors used for the measurement study (left); deployment of the MOX sensors and the reference (right).

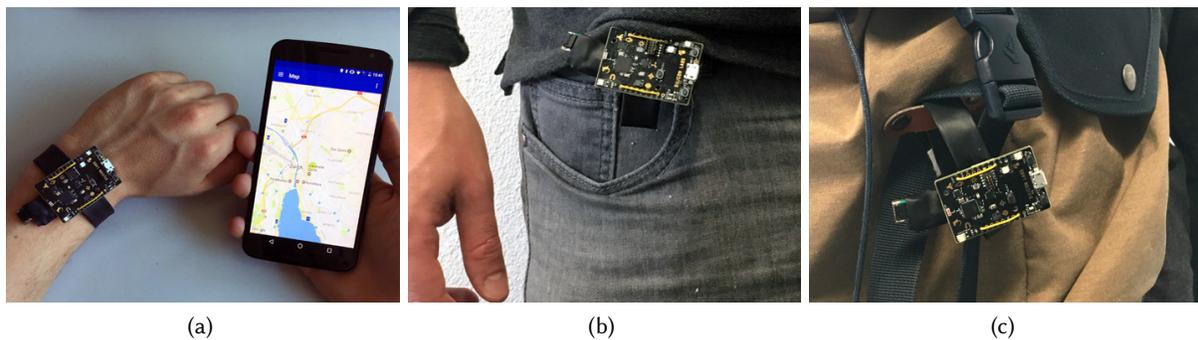


Fig. 2. Different usages of W-Air: (a) worn as wrist-band, (b) attached to a belt and (c) a backpack

Fig. 1. The user is mainly reading or working with a laptop in both environments. In Sec. 7.1 we also investigate the impact of additional human activities. The episodes of measurements with human presence last between 15 and 120 minutes and account for a total of approximately 20 hours for each environment. During situations without human presence the sensors are placed next to each other without any human being present in the close-by vicinity. We only use O<sub>3</sub> reference measurements outdoors due to usually very low and short-lived concentrations indoors [19]. In fact, our O<sub>3</sub> reference sensor always reports 0 ppb when used indoors. The CO<sub>2</sub> references are only used indoors because CO<sub>2</sub> is not considered as a major air pollutant with direct and immediate effect on human beings in outdoor settings [50].

### 3.2 Observations

This subsection presents the key observations through measurements, which motivate the design of W-Air.

There is a clear impact on the measurements of the two MOX sensors caused by human interference (Fig. 3 and Fig. 4). Fig. 3 shows the measurements of the two MOX sensors outdoors when W-Air is used in three different settings, *i.e.*, as a wrist band (Fig. 3a), attached to a belt (Fig. 3b) and to a backpack (Fig. 3c). The same impact is shown when the sensors are placed indoors in Fig. 4. For both environments and all three W-Air usages we observe the same behaviour of the measurements. In the absence of humans, the O<sub>3</sub> measurements (raw and uncalibrated ADC values within [0, 1]) and the VOC measurements (within [0, 1187]) of all the sensors vary moderately. This results is expected because the ground-truth O<sub>3</sub> and CO<sub>2</sub> remain constant within the short time periods. As soon as one person is equipping W-Air the readings of both MOX sensors (O<sub>3</sub> and VOC sensor S<sub>H</sub>) close to the user immediately increase. In fact, in all cases the O<sub>3</sub> values increase between 15% and 40%

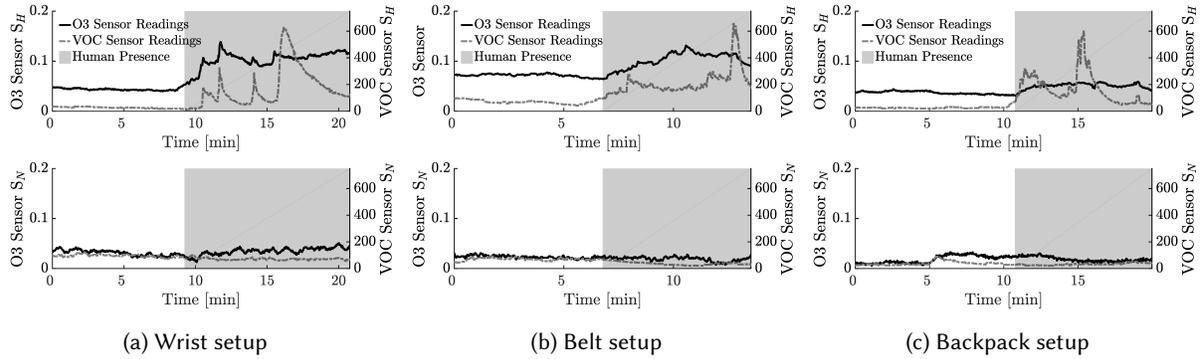


Fig. 3. Outdoor measurements of the human interference when W-Air is worn as (a) a wristband, (b) attached to a belt and (c) attached to a backpack. The  $O_3$  and VOC MOX sensors  $S_H$  are worn by a user and MOX sensors  $S_N$  are placed next to the reference device without any human interference. The ambient  $O_3$  concentration was constant at (a) 30 ppb, (b) 25 ppb and (c) 25 ppb.

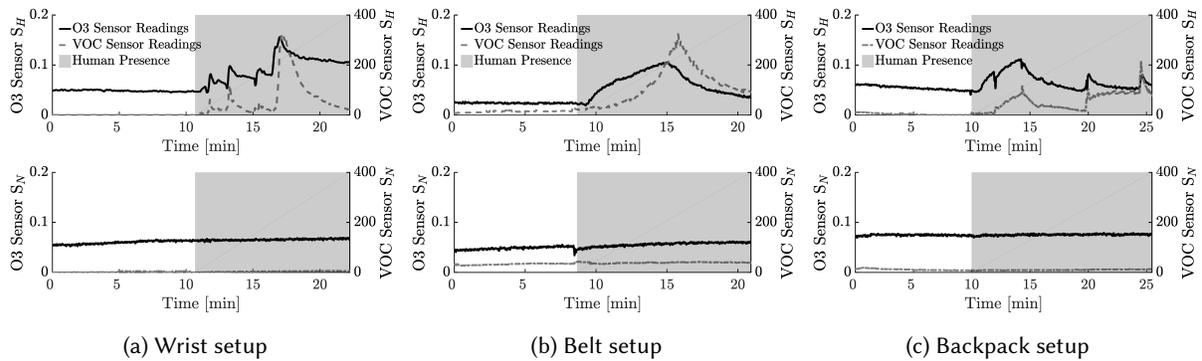


Fig. 4. Indoor measurements of the human interference when W-Air is worn as (a) a wristband, (b) attached to a belt and (c) attached to a backpack. The  $O_3$  and VOC low-cost MOX sensors  $S_H$  are worn by a user and sensors  $S_N$  are placed next to the reference device without any human interference. The ambient  $CO_2$  concentration was constant at (a) 700 ppb, (b) 675 ppb and (c) 700 ppb.

relative to the situation without any human presence. Note an increasing raw value is indicating a decreasing  $O_3$  concentration [51]. The VOC values show peaks over 600 ppb outdoors and over 200 ppb indoors. Yet the usual VOC concentration measured during non human presence situations is below 100 ppb.

One explanation on such abnormal ambient  $O_3$  and VOC measurements may be different VOC emissions:

- **Human skin emissions:** Human skin can emanate different VOCs, both from natural skin oils or from ingredients in cosmetic products [9, 53]. These emissions can further be magnified by their reaction with ambient  $O_3$  [54]. Previous research [9] has identified the depletion of ozone while increasing certain organic components in simulated office environments due to the chemical reactions with skin oils.

- **Textile emissions:** The different skins oils and cosmetics are usually also found in clothing [38, 53], which constitute as a substantial VOC source. Again  $O_3$  can also react with textiles and increases the VOC concentration while being decreased.
- **Human breath emissions:** Exhaled breath of human beings contains various different VOCs [7, 53]. In indoor environments these emissions can have an important impact on the overall air quality.

As a result during situations with human presence we generally observe an (i) increased VOC concentration and a (ii) decreased  $O_3$  concentration. These changes are captured by the two MOX sensors. The  $O_3$  sensor is not only measuring the decreasing  $O_3$  concentration in outdoor environments but also the increasing VOC concentrations indoors where background ozone is usually very low. This is due to the typical cross-sensitivity of MOX sensors. Although the  $O_3$  sensor is designed to be most sensitive to  $O_3$  it is also sensitive to VOCs. Vice-versa the same behavior holds for the VOC sensor. This cross-sensitivity during human presence situations poses a substantial challenge to capture accurate air quality measurements.

The two MOX sensors approximately 2 m away are not affected and remain in the reasonable range, indicating no notable human interference. This is because the convective boundary layer of a standing person in quiet ambient air is within 1m in diameter [12]. Human emissions from more than 1m away can be ignored.

We conclude that state-of-the-art MOX sensors are prone to human interference due to their typical cross-sensitivities to VOCs [51] when used in typical wearable devices. In Sec. 7.1 we also discuss the impact of additional factors on this human interference, such as different human activities or weather situations.

#### 4 EXPLORATORY CALIBRATION OF HUMAN INTERFERENCE

Human interference severely downgrades the accuracy of MOX gas sensors. Therefore, we investigate in this section how we can tackle the human interference problem by exploratory sensor calibration. A large body of existing research [14, 20, 25] shows that calibration of MOX-based air quality sensors is able to provide accurate and stable measurements during situations without human presence. Based on the measurements presented in Sec. 3.1 we investigate in this section the effectiveness of calibration methods during human presence situations.

##### 4.1 Methods

Most of the state-of-the-art calibration methods are either based on linear models, such as *Multiple Linear Regression* (MLS) [25], or non-linear models, such as *Artificial Neural Networks* (ANN) [45]. In general linear calibration models are preferred over non-linear ones due to (i) less vulnerability to over-fitting, (ii) lower computational complexity and (iii) reduced training dataset requirements. Non-linear models on the other hand are more suited for complex calibration problems and can in general, if trained successfully, achieve higher data quality. Therefore, we investigate the performance of MLS and ANNs when applied to our collected data. The ANN is based on a multilayer perceptron using two hidden layers with 100 and 10 neurons respectively.

In order to assess the performance of the models we use different metrics that are widely used in assessing the performance of air quality measurements [45, 47].

**Root-mean-square-error (RMSE).** The RMSE [47] value is a standard metric to assess the calibration error and is defined as follows

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right)^{\frac{1}{2}}, \quad (1)$$

where  $\hat{y}_i$  are calibrated sensor measurements and  $y_i$  ground-truth measurements over a window of  $i = 1, \dots, n$  samples.

Table 2. Calibration results for the two target pollutants applied to a linear calibration model (multiple least squares) with different inputs when sensors are not affected by human interference.

Target	Method	Inputs	RMSE [ppb]	RMSE <sub>σ</sub>	R <sup>2</sup>	β <sub>0</sub> ± t <sub>95%</sub>	β <sub>1</sub> ± t <sub>95%</sub>
O <sub>3</sub>	Linear	{O <sub>3</sub> , T}	7.4	1.5	0.78	0 ± 1	1 ± 0.04
CO <sub>2</sub>	Linear	{VOC, T}	81	0.55	0.88	0 ± 11	1 ± 0.01

**Standardized RMSE (RMSE<sub>σ</sub>).** In order to assess if our calibrated air quality measurements are of sufficient quality for a given application we adopt a standardized RMSE [47] defined as

$$RMSE_{\sigma} = \frac{\left(\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2\right)^{\frac{1}{2}}}{\sigma \cdot \left(\frac{1}{n} \sum_{i=1}^n (y_i)^2\right)^{\frac{1}{2}}} = \frac{RMSE}{\sigma \cdot RMS_y}, \quad (2)$$

where  $RMS_y$  is the root-mean-square value of the ground-truth  $y$  and  $\sigma$  is defined as a relative uncertainty measure for a specific pollutant. For instance, according to the air quality directive by the European Parliament [49],  $\sigma = 0.15$  is required for O<sub>3</sub> measurements. To the best of our knowledge there is no  $\sigma$  defined for CO<sub>2</sub>, for simplicity we assume it is also  $\sigma = 0.15$ . The  $RMSE_{\sigma}$  acts as a statistical parameter to assess the quality of the measurement. If  $RMSE_{\sigma} \leq 1$  the measurements fulfill the air quality directive and can be used for accurate air quality measurements. In the case where  $1 < RMSE_{\sigma} \leq 2$  the measurement quality is not in line with the air quality directive but might still be used for certain applications. For instance, the data quality allows indicative measurements, *i.e.*, assessing the air quality in terms of pollution levels or an air quality index (AQI) [50]. The quality of the measurements is not sufficient for any application if  $RMSE_{\sigma} > 2$ .

**Coefficient of determination R<sup>2</sup>.** The  $R^2 \in [0, 1]$  value is a widely used metric to assess the amount of variance in the calibrated measurements that can be explained by the calibration model. Values close to 1 indicate a well-fitted calibration model.

**Prediction confidence.** A perfect calibration model is able to estimate the exact ground-truth measurements, *i.e.*,  $y_i = \hat{y}_i$ . In reality this is rarely the case and the relationship between ground-truth and estimation can be formulated as  $y_i = \beta_0 + \beta_1 \cdot \hat{y}_i$ , where  $\beta_0 = 0$  and  $\beta_1 = 1$  for a perfect model. A intercept  $\beta_0$  far from 0 indicates that predictions are systematically too low or too high and a slope  $\beta_1$  far from 1 indicates overfitting [46]. We calculate  $\beta_0$  and  $\beta_1$  as well as their T-test 95% confidence intervals by fitting a simple least squares regression on the calibration prediction and the ground-truth measurements. This approach provides a statistical confidence on the model predictions.

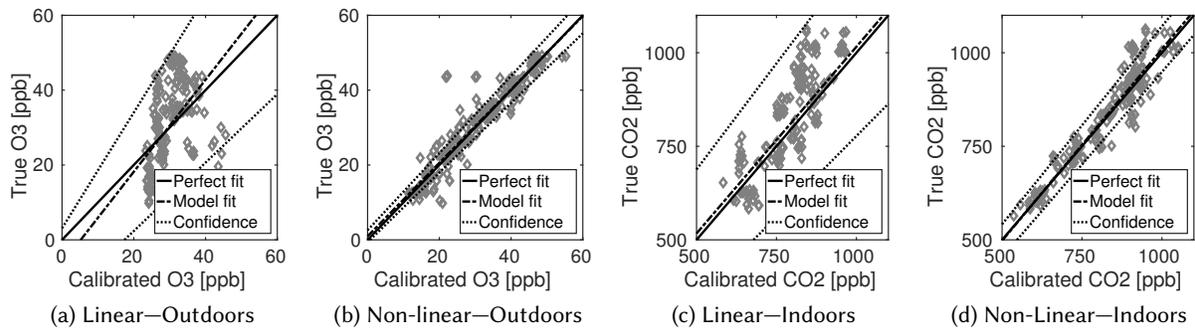
**Cross-validation.** Finally, all the results are based on a 10-fold cross-validation on 2000 samples from our dataset described in Sec. 3.1. As input features we use different combinations of the two MOX sensor readings (O<sub>3</sub> and VOC) and the temperature measurements (T).

## 4.2 Observation

We first obtain a baseline calibration using only data from situations without any human presence. Such a baseline calibration is sometimes provided by the manufacturer for integrated circuit sensor solutions [43]. Table 2 summarizes the average metrics over all 10 folds when we use the linear MLS calibration technique in both environments. Outdoors we predict the O<sub>3</sub> concentration using the O<sub>3</sub> MOX and temperature sensor and indoors we predict the CO<sub>2</sub> concentration using the VOC MOX and temperature sensor. We observe that already a linear model achieves accurate results in both environments. The  $RMSE_{\sigma}$  of approximately 1.5 for the O<sub>3</sub> calibration indicates that the measurements can be used for indicative air quality results. Recent studies also

Table 3. Calibration results for the two target pollutants applied to different methods and inputs during human interference.

Target	Method	Inputs	RMSE [ppb]	RMSE <sub>σ</sub>	R <sup>2</sup>	$\beta_0 \pm t_{95\%}$	$\beta_1 \pm t_{95\%}$
O <sub>3</sub>	Linear	{O <sub>3</sub> , T}	10.8	2.24	0.15	0 ± 9.9	1 ± 0.33
O <sub>3</sub>	Linear	{O <sub>3</sub> , VOC, T}	10.7	2.23	0.16	0.16 ± 9.8	0.99 ± 0.32
O <sub>3</sub>	Non-Linear	{O <sub>3</sub> , T}	3.9	0.81	0.89	0.7 ± 1.5	0.98 ± 0.05
O <sub>3</sub>	Non-Linear	{O <sub>3</sub> , VOC, T}	3.5	0.72	0.91	0.8 ± 1.3	0.98 ± 0.04
CO <sub>2</sub>	Linear	{VOC, T}	182	1.56	0.1	-395 ± 435	1.5 ± 0.6
CO <sub>2</sub>	Linear	{VOC, O <sub>3</sub> , T}	135	1.17	0.49	1.1 ± 108	0.99 ± 0.14
CO <sub>2</sub>	Non-Linear	{VOC, T}	68	0.6	0.86	34 ± 39	0.95 ± 0.05
CO <sub>2</sub>	Non-Linear	{VOC, O <sub>3</sub> , T}	44	0.38	0.94	16 ± 24	0.98 ± 0.03

Fig. 5. Exploratory sensor calibration during human interference using O<sub>3</sub>, VOC and temperature sensor readings with: (a) linear calibration (multiple least squares [25]) outdoors; (b) non-linear calibration (artificial neural network [45]) outdoors; (c) linear calibration indoors; (d) non-linear calibration indoors.

show that the quality of low-cost O<sub>3</sub> sensor measurements can be improved to fulfill the air quality directive using more complex calibration models [45]. The CO<sub>2</sub> calibration performs even better with a RMSE<sub>σ</sub> of 0.55. These results have already been shown before in related work [20, 25, 45].

In a next step we investigate the calibration during human presence situations. The coherent changes in the measurements of co-located O<sub>3</sub> and VOC sensors when there is human interference lead us to explore joint calibration of O<sub>3</sub> and VOC measurements. Since we aim to perform calibration on resource-constrained wearable and mobile devices, a natural question arises *which sensor measurements* we use as inputs and what *model complexity*, e.g., linear versus non-linear, is necessary.

Table 3 presents the result of the calibration when using data during human presence situations. We compare the non-linear ANN calibration to the linear MLS one with different inputs. We observe that the linear model for both indoor and outdoor environments performs notably worse than during situations without human interference. The RMSE<sub>σ</sub> is in all cases above 1 and therefore not satisfying the data quality goals. Further, the R<sup>2</sup> values are in all cases below 0.5, indicating a poorly fitted model. This result is reflected in the scatter plots in Fig. 5a and Fig. 5c where we plot the predictions of the testing dataset of the 10th fold against the corresponding ground-truth. The X-axis and the Y-axis denote the calibrated and the ground-truth measurements respectively. The fitted model exhibits poor confidence, i.e., the regression line defined by  $\beta_0 \pm t_{95\%}$  and  $\beta_1 \pm t_{95\%}$ . This result changes if we use a non-linear ANN calibration model. All metrics drastically improve and perform best when

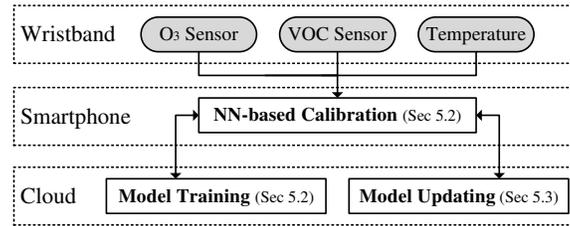


Fig. 6. Work flow of W-Air. A sensor array consisting of a low-cost  $O_3$ , VOC and temperature sensor is utilized to calibrate  $O_3$  and  $CO_2$  (estimated by VOC readings) concentrations. Calibration is based on neural networks and performed on a smartphone. The cloud server performs calibration parameter training and updating for different individuals to provide personalized calibration parameters and adapt to new environments.

we use all three sensor measurements as inputs. The  $RMSE_{\sigma}$  is below 1 and, thus, implying sufficient data quality. In Fig. 5b and Fig. 5d we also observe the improvement of the prediction confidence. The fitted model is close to the perfect fit with a significantly higher confidence compared to the linear model.

We conclude, that a non-linear neural network is able to compensate for the human interference and recover the true  $O_3$  and  $CO_2$  concentrations with low error and high confidence.

**Summary.** Emissions of human beings impose non-linear interference on ambient  $O_3$  and VOC concentrations. The non-linearity might come from the fact that low-cost MOX sensors are optimized to work linearly within a limited concentration range. Sensors designed and calibrated for lightly polluted areas may suffer severe non-linearity in the high concentration range (due to human interference). At minimal, a non-linear calibration scheme that combines raw  $O_3$  measurements, VOC measurements and environmental factors is indispensable to accurately calibrate the  $O_3$  readings outdoors and  $CO_2$  readings indoors. Despite previous work on non-linear calibration [4, 45] for static sensor arrays, an effective and efficient sensor array calibration for wearables is missing, which motivates the design of W-Air.

## 5 SYSTEM DESIGN

This section first presents the overview of W-Air and then elaborates on the calibration scheme with a focus on the techniques to improve the usability. Finally, we present the implementation of W-Air in Sec. 5.4.

### 5.1 Overview

As illustrated in Fig. 6, W-Air consists of (i) wristbands integrated with COTS air pollutant sensors; (ii) smartphones; and (iii) a cloud server. Users wear W-Air on wristbands, attach it to belts or to backpacks to measure raw  $O_3$  and VOC concentrations as well as environmental factors such as temperature. They also carry smartphones, which calibrates the raw measurements to eliminate the human interference and other environmental factors to output accurate ambient  $O_3$  concentration (outdoor) and  $CO_2$  concentration (indoor). The cloud server communicates with W-Air clients and is responsible for calibration model training and updating.

There are two major functional components in W-Air: (i) neural network based calibration, and (ii) calibration training and updating. The core technique to filter the human interference and environmental factors (e.g., temperature) is the **neural network (NN) based calibration scheme** (Sec. 5.2). We explicitly distinguish the cases with and without the human interference for more effective calibration and design two separate neural networks for outdoor  $O_3$  calibration and indoor  $CO_2$  concentration calibration, respectively. For easy maintenance and usage, we apply a NN architecture with shared hidden layers for **model training** (Sec. 5.2) with fewer

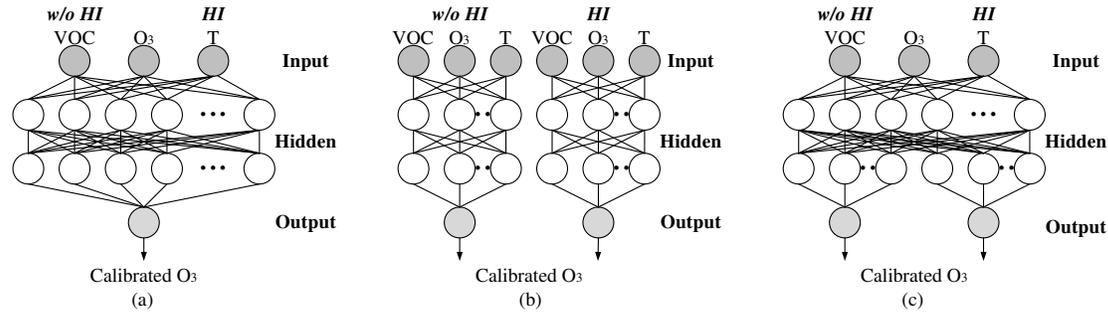


Fig. 7. Architectures of neural networks (NNs) for  $O_3$  calibration: (a) a single NN; (b) two independent NNs; (c) two NNs with shared hidden layers. Here HI is short for human interference.

samples than traditional NN architectures and utilize **semi-supervised updating** (Sec. 5.3) to adjust model parameters with little human intervention.

W-Air stores two neural network models to calibrate  $O_3$  and  $CO_2$ , respectively, but only runs one to output accurate  $O_3$  concentration *outdoors* and  $CO_2$  concentration *indoors*. This is because  $O_3$  is one of the major air pollutant outdoors and the  $O_3$  concentration is expected to be zero indoors [52]. Conversely,  $CO_2$  is an important indicator for indoor air quality [44]. To detect whether a W-Air user is indoor or outdoor, any indoor/outdoor detection scheme applies [37, 57].

For brevity, the rest of this section mainly focuses on  $O_3$  calibration. Model structure, training and updating for  $CO_2$  calibration all follow the same principle and only differ in specific parameters.

## 5.2 Calibration Methods

As depicted in Sec. 3.2, a non-linear model *e.g.*, a neural network (NN) that inputs  $O_3$ , VOC and temperature measurements is sufficient for accurate compensation of human interference. However, we argue that a naive neural network is challenging to apply for the following reasons. (i) A complete calibration scheme should operate both *with* and *without* human interference. (ii) An easy-to-use calibration scheme should involve little training effort. (iii) The calibration performance with and without human interference should be relatively accurate despite imbalanced training dataset sizes for the two cases. In the following, we assume the labels of the two cases (with or without human interference) are known a priori, and discuss the potential for automatic detection of human interference in Sec. 7.2.

There are three candidate NN architectures for calibration. Fig. 7 illustrates the architectures for  $O_3$  calibration. The same architectures also apply for  $CO_2$  calibration. An intuitive proposal is to adopt a unified NN that calibrates  $O_3$  measurements for both cases, with human interference and without human interference (Fig. 7(a)). However, a unified, fully-connected NN model is vulnerable to imbalanced training data.

Imbalanced training data are common in practice because most sensor manufactures only perform generic calibration in laboratories. That is, the sensors are calibrated without human interference. Field calibration is necessary before the sensors can achieve the claimed accuracies [14, 20, 25]. However, end users are reluctant to collect sufficient measurements for field calibration (with human interference in our context). In this case, W-Air needs to train calibration parameters with limited samples with human interference.

Since the  $O_3$  measurements behave differently with and without human interference, the calibration parameters for the two cases also differ notably. With an extremely imbalanced training dataset, a unified NN model tends to train the calibration parameters for human interference largely on the samples collected without human

interference. Alternatively, we can create two independent NNs for the cases with human interference and without human interference and train each NN separately (Fig. 7(b)). Nevertheless, this scheme may involve substantial training data for the two cases since the samples for human interference are of no help to train the NN for the case without human interference. Note that the calibration models for human interference and without human interference may share certain common features. Hence a more efficient architecture is to allow shared hidden layers for the two cases (Fig. 7(c)). By enabling data sharing between the two cases, samples for human interference will contribute to extracting the common features useful for calibration when there is no human interference. Consequently, the amount of training data necessary to train the calibration models for both cases will decrease. We evaluate the efficiency of the three NN architectures in detail in Sec. 6.2.

By default, we assign two hidden layers for architecture (a) and (b) with 100 neurons in the first layer and 10 neurons in the second layer. Architecture (c) is composed of 1 shared layer with 100 neurons and two additional separate layers with 10 neurons each. The cost functions of all architectures are to minimize the mean-squared-error (MSE) between the calibrated and ground-truth  $O_3$  readings. We observed through extensive testing that these architecture setups achieve accurate and robust output values. Especially less than 10 neurons in the second layer has a negative effect on the calibration error. Also, we chose a relatively flat architecture because it is easier to train and less prone to overfitting than architectures with more hidden layers and neurons. We apply the same procedure to select the hyper-parameters in the NN for VOC (or equivalently  $CO_2$ ) calibration.

### 5.3 Parameter Updating

For accurate calibration, it is essential to train the NNs on abundant and diverse data covering different environments. The optimal calibration parameters may differ for different environmental conditions (e.g., sunny versus rainy weather). Additionally, the optimal calibration parameters for each user can also differ because the living environments of users may notably differ. Therefore it is important to tune the NNs to adapt to the actual target environments. Given sufficient *labelled* measurements from the target environments, the calibration parameters can be trivially updated by retraining the NNs adding the newly labelled measurements. Here a *labelled* measurement refers to a tuple of VOC,  $O_3$  and temperature together with the true  $O_3$  value (i.e., the label). However, it is cumbersome, if not impossible for users to label the new  $O_3$  measurements collected from his/her own living environments for W-Air to learn and update the calibration parameters. This is because the calibration is a regression problem, and users need to carry a highly reliable reference  $O_3$  sensor to label the ground-truth  $O_3$  concentration, which is impractical in our application scenarios.

To allow W-Air to adapt to new environments, we harness the paradigm of semi-supervised learning, where “unlabelled” (without ground-truth) data are used to improve the accuracy of learning models [59]. In W-Air, we apply COREG [58], a co-training style semi-supervised regression framework to update the calibration models with unlabelled gas measurements. Co-training operates by running two regressors iteratively [59]. The two regressors assign pseudo-labels to each unlabelled sample, and the pseudo-label with higher confidence is used to retrain and improve the performances of both regressors. COREG [58] utilizes two k Nearest Neighbour (kNN) regressors for co-training. The effectiveness of the co-training is fulfilled by using different distance metrics and the number of nearest neighbours (i.e., k) for the two kNNs. After assigning the unlabelled samples using the kNNs, any regressor can be constructed based on the extended (containing both labelled and pseudo-labelled samples) training set to improve regression accuracy.

Fig. 8 illustrates the work flow of COREG in the context of W-Air. Initially, two kNNs ( $kNN_1$  and  $kNN_2$ ) are trained using the labelled sets  $L_1$  and  $L_2$ , respectively. In each iteration, an unlabelled sample ( $\langle \text{VOC}, O_3, T \rangle$ ) is assigned a pseudo-label (calibrated  $O_3$ ) by  $kNN_1$  ( $kNN_2$ ), and the most confident pseudo-labelled sample (calculated based on MSE, see [58] for details) is moved to  $L_2$  ( $L_1$ ) to retrain  $kNN_2$  ( $kNN_1$ ). The outputs of COREG are two extended training sets  $L_1$  and  $L_2$ , where two NNs are constructed. A new testing sample will be calibrated

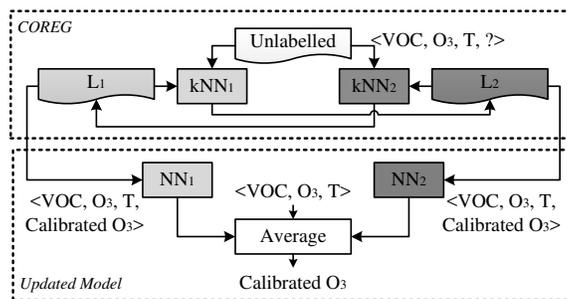


Fig. 8. An illustration of the COREG work flow for  $O_3$  calibration in W-Air.

by the two NNs, and the final calibration result is the average of the outputs of the two NNs. Note that the co-training framework is applied for W-Air to adapt to different environments and is performed for both human interference and without human interference.

#### 5.4 Implementation

This subsection describes the implementation of W-Air.

**5.4.1 Hardware.** We integrate two mainstream COTS MOX sensors, a *MICS-OZ-47* and a *CCS811*, into a *Thunderboard Sense* development platform (see Fig. 2). Note that W-Air serves as a proof-of-concept for demonstrating and eliminating the human interference problem. More compact mechanical designs are out of the scope of this paper. Further, we use the on-board *Si7210* temperature and relative humidity sensors. By default, the MOX gas sensors and the temperature sensor sample at 2 Hz when turned on. The *Thunderboard Sense* provides a GATT-Server data architecture to provide raw sensor measurements via Bluetooth LE.

We use a *Motorola Nexus 6* smartphone featuring a 2.7 GHz quad-core CPU, 3 GB of RAM and a 3220 mAh battery, running Android 7.0 OS for calibration. The smartphone communicates with the wristband via Bluetooth. A *Lenovo Thinkpad T440p* featuring a 2.5 GHz octa-core CPU, 16 GB RAM and running Ubuntu 16.04 serves as the cloud server for calibration model training and updating.

**5.4.2 Software.** We apply an indoor/outdoor inference scheme similar to [57] to trigger the corresponding neural network for outdoor  $O_3$  calibration and indoor  $CO_2$  calibration. We implement the indoor/outdoor inference and the calibration scheme in *python*. The neural networks are implemented using *Tensorflow*, an open-source machine learning library by Google [11]. We train the indoor/outdoor classifier and the calibration model on the cloud server. The final classifiers and calibration networks are exported and integrated into an Android application. Tensorflow facilitates updating of the neural networks after the initial application installation. Thus, re-training can be done on the cloud server and seamlessly exchanged in the application of a user without recompilation.

## 6 EVALUATION

In this section we thoroughly evaluate the overall calibration performance of our NN-based approach (Sec. 6.1), the benefits of using a NN structure with a shared layer (Sec. 6.2) and the effectiveness of semi-supervised model updating (Sec. 6.3).

Table 4. Overall calibration errors denoted as RMSE ( $RMSE_{\sigma}$ ) of a linear baseline calibration [25] and our NN-based approach.

Environment	Test Case	Baseline	NN-based
Outdoor ( $O_3$ )	No Human Interference	10 ppb (2.0)	4.3 ppb (0.86)
	With Human Interference	16.8 ppb (3.7)	4.3 ppb (0.94)
Indoor ( $CO_2$ )	No Human Interference	177 ppb (1.5)	64 ppb (0.57)
	With Human Interference	325 ppb (2.3)	38 ppb (0.29)

## 6.1 Overall Calibration Performance

We first show the overall performance of our NN-based calibration and its advantage over a naive baseline, *i.e.*, a typical calibration provided by sensor manufacturers [43]. The dataset is based on the measurements described in Sec. 3.1, *i.e.*, we calibrate the W-Air measurements to the ground-truth measurements provided by the two reference sensors. The RMSE and its standardized version  $RMSE_{\sigma}$  as described in Sec. 4.1 are used to assess the performance of the calibration. As a baseline we construct a linear model that calibrates the raw  $O_3$  (VOC) sensor and temperature measurements to the actual ground-truth  $O_3$  ( $CO_2$ ) concentration when there is no human interference. This baseline is first trained on 1000 samples when there is no human presence. Finally the baseline is tested on 10000 samples each for both cases, *i.e.*, with and without human interference, and the errors shown in Table 4. The testing dataset includes measurements from various locations and times as described in Sec. 3. The baseline performs well for measurements without human interference with an error that is acceptable for indicative measurements. However during situations with human interference the error is notably higher and the  $RMSE_{\sigma}$  is larger than 2 for both  $CO_2$  and  $O_3$ . Thus, these calibrated measurements are not usable for any conclusions about the actual air pollution. When applying our non-linear NN-based calibration (shared layer architecture trained with 1000 samples for each case) we achieve the best performance for both cases and both pollutants with a RMSE of 4.3 ppb for  $O_3$  measurements and 64 ppb for  $CO_2$  measurements. In addition, we achieve a  $RMSE_{\sigma} < 1$  for both pollutants.

In conclusion, our NN-based approach is able to infer the true ambient  $O_3$  and  $CO_2$  concentration during situations with and without human interference with an accuracy that is sufficient for personal air pollution monitoring.

## 6.2 Effectiveness of NN Architecture

**6.2.1 Settings.** In this section we evaluate the effectiveness of our NN-based calibration. We make again use of the measurements described in Sec. 3.1. The following results are based on a 10-fold cross-validation and evaluated on the  $RMSE_{\sigma}$ . We evaluate the calibration accuracy of the three NN architectures (see Sec. 5.2) using both balanced and imbalanced training sets in sequel.

**6.2.2 Calibration Performance Using Balanced Training Sets.** We first investigate the calibration performance if we have the equal amount of training samples (balanced training sets) for both cases, *i.e.*, case 1 with human interference and case 2 without. Fig. 9a and Fig. 9c depicts the calibration performance averaged for human interference and without human interference using balanced training sets for  $O_3$  outdoors and  $CO_2$  indoors respectively. As expected with more training samples, the calibration errors for all the three NN architectures gradually decreases. We observe that two independent NNs perform in average better than a single unified NN when trained with few samples, indicating the necessity to distinguish the two cases. By allowing data sharing between the two cases, *i.e.*, using a NN with a shared layer, the calibration error can be additionally decreased. W-Air achieves an up to 10 % lower error for both  $O_3$  and for  $CO_2$  than the two independent NNs using the same

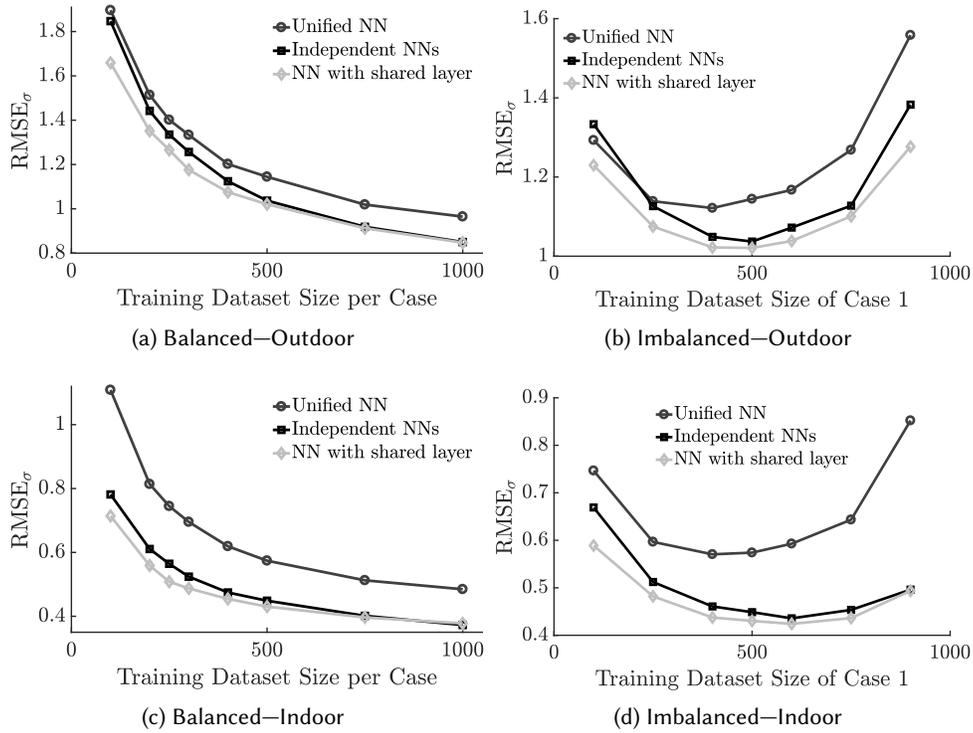


Fig. 9. Average calibration errors denoted as  $RMSE_{\sigma}$  for both human interference and without human interference using (a) increasing amounts of balanced training samples and (b) imbalanced training sets in an outdoor environment. The same evaluation for indoor environments is presented in (c) and (d) respectively.

amount of training data. For training datasets with more than 500 samples both structures perform equally. Thus, our NN approach with a shared layer is particularly helpful when available training data is limited.

**6.2.3 Calibration Performance Using Imbalanced Training Sets.** Due to the diversity in user lifestyles, users are likely to collect different proportions of measurements with and without human interference. Note that we do not assume users always wear W-Air to take measurements. Users may utilize W-Air as a static air pollution sensor when not wearing it. Hence it is crucial that W-Air still works both with and without human interference, even when trained by imbalanced data.

Fig. 9b and Fig. 9d plot the average calibration errors using a fixed 1000 training samples with varying numbers of measurements with human interference for outdoors and indoors respectively. We choose a total of 1000 samples for training because roughly 500 samples are needed to successfully train a well-performing NN for one case (see Fig. 9a and Fig. 9c). The results show that the shared layer architecture always yields lower calibration errors, up to 8% outdoors and 12% indoors. This finding highlights the benefit of using a shared layer. We can exploit training samples from one case to train the other case due to common features modeled within the shared layer.

Fig. 10 further depicts the calibration errors using imbalanced training sets for the two cases separately. We again observe that the shared layer architecture is usually most efficient when using small amounts of training

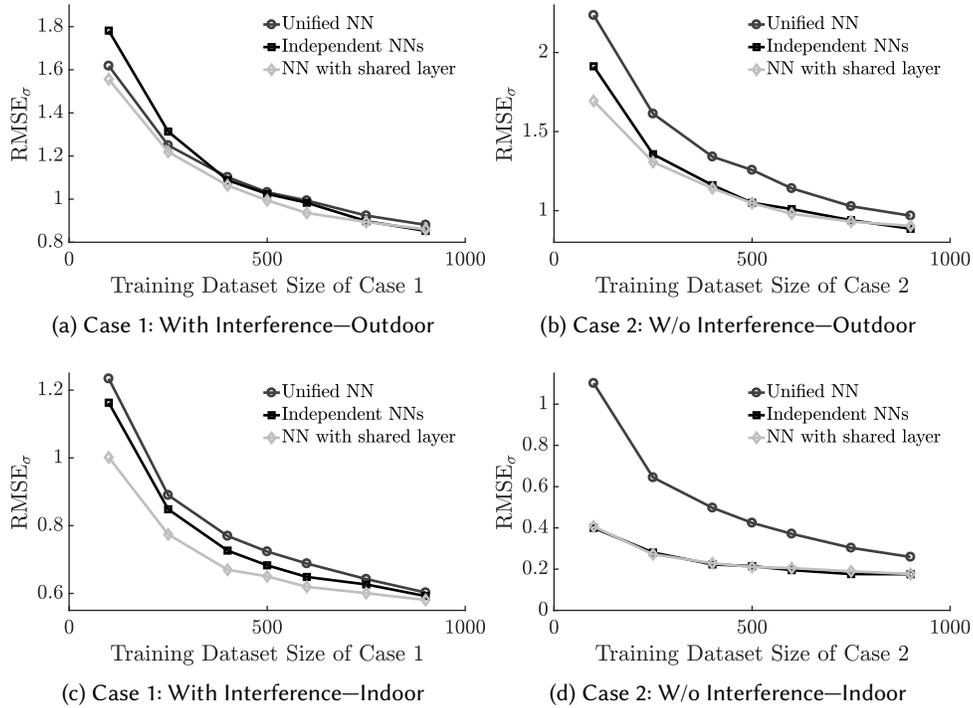


Fig. 10. Calibration errors with (a) human interference and (b) without human interference using imbalanced training sets for outdoor  $O_3$  measurements. The same evaluation for indoor  $CO_2$  measurements is presented in (c) and (d) respectively.

Table 5. Calibration errors denoted as RMSE (RMSE<sub>σ</sub>) for different users.

	User 1	User 2	User 3	User 4	User 5
$O_3$ outdoors	6 ppb (0.83)	4.5 ppb (0.81)	14.3 ppb (3.8)	12 ppb (2.85)	7.3 ppb (1.86)
$CO_2$ indoors	28.8 ppb (0.29)	115 ppb (1.5)	32 ppb (0.37)	40 ppb (0.47)	38 ppb (0.41)

data. With larger amounts of training data the three different NN architectures perform similar. An interesting observation is the almost identical performance of the shared layer and the independent NN architecture for case 2 indoors. Already a few samples suffice to provide accurate data with an  $RMSE_{\sigma} < 0.5$ . This is due to the strong correlation between the VOC measurements and the ground-truth  $CO_2$  concentration during non-human interference situations.

**6.2.4 Robustness to Different Users.** The effect of human interference depends on the amount of VOC a user emits. Thus, we investigate the calibration accuracy of different users. Five different users, 4 male (user 1 to 4) and 1 female, additionally recorded between 10 and 15 min of data each outdoors and indoors. The users are wearing W-Air on their wrist while standing or sitting close to a reference sensor as illustrated in Fig. 1. We train a calibration NN with the measurements from Sec. 3 recorded by user 1 with 1000 samples for each case and finally test on all the data from the four remaining users. Table 5 summarizes the resulting errors for each user. For outdoor measurements we observe notable differences between the users. Especially the measurements of

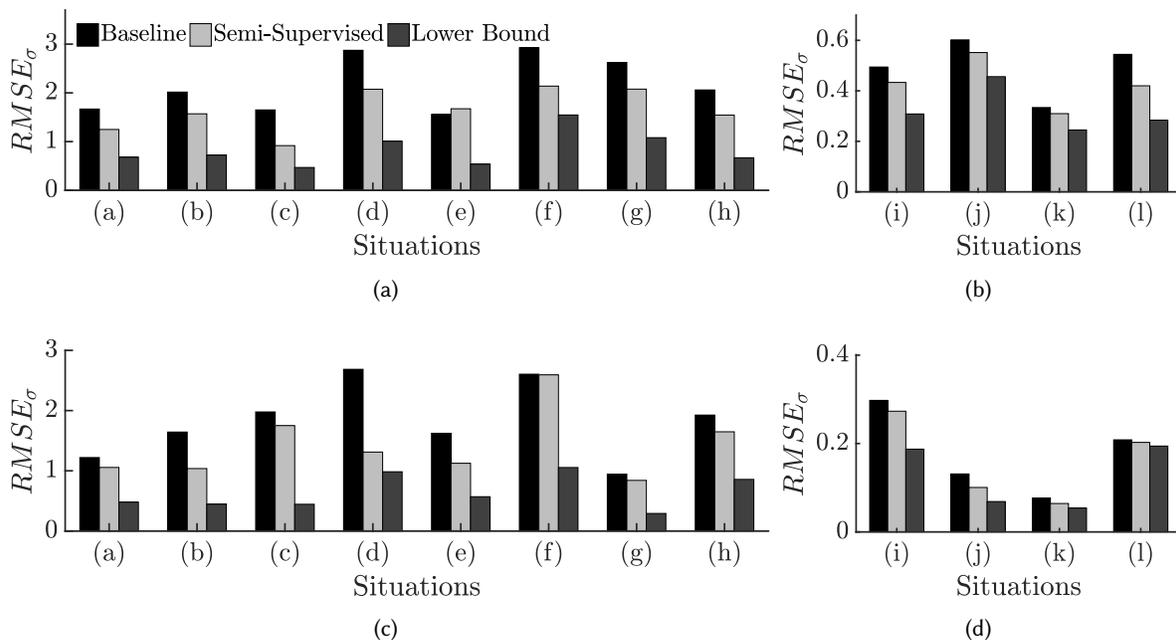


Fig. 11. Semi-supervised learning performance for situations (a) to (l) for case 1 with human interference outdoors (Fig. 11a) and indoors (Fig. 11b) and case 2 without human interference outdoors (Fig. 11c) and indoors (Fig. 11d).

user 3 and 4 suffer from a substantial error that makes it impossible to draw reliable conclusions about their personal exposure to air pollution. A similar behavior we observe indoors for user 2. While all other users achieve accurate results, the results of user 2 suffer from a substantial error that is almost three times higher than the one of the other users. The reason for this behavior are notable differences of the measurements between the different users. This indicates that the human interference can differ between multiple users. We believe that the individual calibration accuracy can be improved by performing per-person training for each user.

### 6.3 Effectiveness of Semi-supervised Updating

**6.3.1 Settings.** The two NNs (*i.e.*,  $NN_1$  and  $NN_2$  in Fig. 8) are based on the shared layer architecture (c) with 100 and 10 neurons each in the hidden layers as used before. For each environment, *i.e.*, indoor and outdoor, we use separate NNs. In order to evaluate the performance of the semi-supervised updating on adapting to new situations, we split our dataset into 12 different distinct situations. Outdoors we distinguish eight different situations based on time and weather conditions because these two factors are the main influences that affect the ambient  $O_3$  concentration. For the remaining 4 indoor situations we distinguish between different air quality levels based on the  $CO_2$  concentration. Table 6 summarizes the different conditions for the twelve situations (a) to (l). We apply a leave-one-out validation with these situations by performing supervised learning on seven outdoor situations. Semi-supervised learning and testing is performed on the remaining situation for all eight possible combinations for the outdoor situations. Equally we apply the leave-one-out validation with the four indoor situations, but added 10% of data from the left-out situation to the training data. This approach shows

Table 6. Different situations used for semi-supervised updating with their relative improvement compared to a baseline and the average number of unlabelled samples added based on the COREG results.

Situation (outdoor)	Time, Weather	Improvement		# Samples added	
		Case 1	Case 2	Case 1	Case 2
(a)	April, afternoon, rain	25.1%	13.5%	58	47
(b)	April, night, cloudy	22.1%	36.1%	68	50
(c)	April, afternoon, sunny and windy	44.4%	11.5%	86	2
(d)	April, morning, sunny	27.8%	51.2%	20	49
(e)	April, morning, light rain	-7.5%	30.7%	25	42
(f)	June, afternoon, sunny and hot	26.9%	0.5%	96	20
(g)	June, night, sunny and hot	20.1%	11.0%	56	53
(h)	August, afternoon, hot and windy	25.4%	14.4%	55	57
<hr/>					
Situation (indoor)	CO <sub>2</sub> concentration				
(i)	low (400-600 ppb)	12.3%	8.2%	96	93
(j)	low-medium (600-750 ppb)	23.3%	3.1%	85	86
(k)	medium (750-850 ppb)	7.2%	16.4%	90	102
(l)	high (850-1000 ppb)	22.9%	2.8%	94	90
Average		19.6%	18.4%	69	58

that the COREG algorithm is able to improve a model that is trained on a dataset with lack of data for certain ground-truth data ranges.

We compare the performance of our semi-supervised learning to a baseline approach, *i.e.*, the testing error using only the supervised model. Further, we compute a lower bound for the error by adding the unlabelled data to the training set including the true labels. We use 750 samples for each case, with and without human interference, as labelled data for training, and a pool of 500 unlabelled samples for the semi-supervised COREG algorithm. The error is tested on 300 different samples of testing data. The evaluation is finally performed on 10 different pools of labelled and unlabelled data.

**6.3.2 Performance.** Semi-supervised learning is able to improve the calibration error in all but one situation. Fig. 11a and Fig. 11c shows the  $RMSE_{\sigma}$  for the eight outdoor situations for case 1, with human interference, and case 2, without human interference, respectively. The same results are shown in Fig. 11b and Fig. 11d for the four indoor situations. The improvement is between 0.5% and 51% and in average about 19.6% and 18.4% for case 1 and case 2 respectively, as summarized in Table 6. Only in situation (e) the semi-supervised updating resulted in a degradation of the accuracy of 7.5% for the case with human interference. The reason might be that this situation is too different from the other situations in terms of measurements. The COREG was in fact able to label only 25 unlabelled samples, which is notably below the average of 69 unlabelled samples. Further, we see that for certain situations the lower error bound is significantly lower than the error of the semi-supervised approach. Especially in outdoor situations we observe the need to adapt our calibration model to new situations. For indoor measurements we observe smaller improvements than for outdoor measurements, in particular for case 2 without human interference. This is due to already relatively well performing calibration model before the semi-supervised update test, *i.e.*, the difference of the baseline and the lower bound are notably smaller compared to the outdoor situations.

It is in general challenging to adapt a model to an unfamiliar situation. Especially in outdoor situations the model can be further improved. Thus, it will be important to also exploit supervised calibration methods, *e.g.*, by exploiting opportunistic calibration with other devices [26, 40] or static reference sensors [13, 14]. We believe

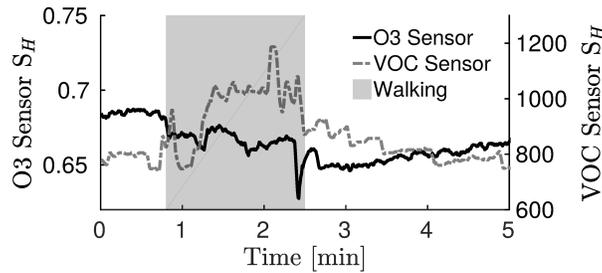


Fig. 12. VOC and  $O_3$  MOX sensor measurements of W-Air attached to a backpack of a user standing and walking indoors. As soon as the user starts to walk we observe an increase of the VOC sensor signal and a decrease of the  $O_3$  sensor signal, indicating an increased intensity of the interference problem.

that a combination of semi-supervised and supervised calibration methods will improve the overall accuracy of W-Air.

## 7 DISCUSSIONS AND FUTURE WORKS

This section discusses the limitations and future improvements of W-Air.

### 7.1 Impact of User Context

**7.1.1 Environmental Conditions.** Environmental conditions have not only an effect on the air pollution but also on the intensity of the human interference. Especially for different weather conditions in outdoor environments we observed different human interference behaviors. For instance, our W-Air device measured an average VOC concentrations of 130 ppb during human interference situations in July and only 35 ppb in October. While the average ground-truth  $O_3$  concentration during both months barely differed with 44 ppb in July and 42 ppb in October, the weather conditions showed notable differences. The temperature was around  $28^\circ C$  in July without any rain and  $18^\circ C$  in October with multiple rain periods.

This dependency of the human interference on environmental conditions poses a substantial challenge for our W-Air system. We need to train the calibration model with data from various environmental conditions in order to provide accurate measurements during these conditions. This is possible by exploiting semi-supervised updating as we show in Sec. 6.3. To further improve the measurement accuracy it will also be important to incorporate supervised techniques such as opportunistic calibration [26, 40].

**7.1.2 Activities.** The current activity of a user is known to be an interfering factor for high data quality in wearable sensing [13, 24]. In order to exclude those influences in our evaluation we only considered situations with the user either sitting, standing or doing office-work. Other daily activities, such as walking, running or exercising can in fact have an additional impact on the sensor readings, *i.e.*, an increased intensity of the human interference problem. In Fig. 12 we investigate the impact of walking on the sensor data. We observe that as soon as the user starts to walk the VOC sensor measurements increase and the  $O_3$  sensor measurements decrease. As we show in the measurement study in Fig. 3, this is the typical behavior of the human interference problem. The impact of walking diminishes again as soon as the user stops walking around 2:30 min. The reason for this behavior might be the consequence of the increased air flow over the sensing layer and, thus, the two gas sensors react differently. We also observe this behavior when using W-Air in the wrist and belt setting, as well as in outdoor environments.

Consequently not only the VOC emissions of a user but also its different activities, especially mobility [13], can interfere with the sensor readings. In a first step of possible future work it will be important to thoroughly study the impact of human activities on the air quality measurements of our W-Air system. In a second step it is equally important to adapt and expand our sensor calibration by also incorporating different human activities.

## 7.2 Automatic Detection of Human Interference

The distinctive characteristics of the human interference (Sec. 3.2) lead us to distinguish between the cases with and without human interference. In this work, we assume the knowledge of whether there is human interference is known in advance. Autonomous human interference detection is feasible by leveraging additional sensors of the wearable device. We can use accelerometers to detect movements or a switch that closes a circuit when one clips on the wristband.

Some pioneer studies have integrated proximity sensors into watchbands for wrist gesture input [10], which can also be applied to detect whether a wristband or a smart-watch is worn by the user. Autonomous human interference detection will further reduce the overhead for users to label new measurements to update the calibration parameters.

## 7.3 System Performance

Energy and delay are two crucial factors for practical wearable systems. The most energy-hungry components of W-Air are the two MOX based gas sensors. We measure an average current draw of 50 mA when turning both gas sensors on. Continuously powering the gas sensors will drain a standard 250 mAh coin cell battery within 5 hours and, thus, would limit usability to a great extent. One future direction to improve the energy efficiency of W-Air is to duty-cycle the gas sensors in undesired situations. For instance, the average current draw of the VOC sensor integrated in W-Air drops to 0.7 mA with a duty-cycle of 60 s. However, long duty-cycles also cause slower response time and lower sensitivity of the sensor [2].

In terms of delay, the most computational and time intensive task of W-Air is the NN-based calibration. The majority of the delay is induced by the feature extraction, namely a 3 sec smoothing window for the sensor measurements. Running the neural network to output the final calibrated concentrations takes negligible time (around 5 ms). However, there is a delay of around one minute when first turning on the gas sensors before they work in a stable state. During this warm-up time W-Air is not providing accurate measurements. In a duty-cycling scenario the trade-off between energy efficiency and warm-up delay needs to be carefully evaluated. We also envision the one-shot delay due to the warm-up time of the gas sensors will be notably reduced by the rapid development of sensor technologies.

## 8 CONCLUSION

In this work, we propose W-Air, an accurate personal multi-pollutant monitoring platform for wearable devices. We identify the human interference problem when integrating low-cost MOX gas sensors into wearable platforms such as wrist-worn, attached to belts or backpacks, which is largely overlooked in previous research. We propose a partitioned neural network based sensor array calibration scheme to eliminate the non-linear human interference on ambient outdoor O<sub>3</sub> measurements and indoor CO<sub>2</sub> measurements. The architecture of the calibration model is carefully devised to reduce training efforts and facilitate parameter updating with little human intervene. We prototype W-Air on a wristband with low-cost COTS gas sensors. Evaluations show that W-Air is able to yield accurate ambient O<sub>3</sub> and CO<sub>2</sub> measurements whether there is human interference. It can also learn from unlabelled measurements and adapt to unfamiliar circumstances by exploiting semi-supervised learning.

Future directions include (i) more thorough investigations how user context and activities affect the human interference problem, (ii) combining W-Air with state-of-the-art health and fitness trackers to explore correlations between air pollution and wellbeing of a user and (iii) large-scale deployments for extensive user-studies.

## ACKNOWLEDGMENTS

We would like to thank Florian Zaruba and Thomas Brunner for their valuable contributions. This work was funded by the Swiss National Science Foundation (SNSF) under the FLAG-ERA CONVERGENCE project.

## REFERENCES

- [1] Aeroqual. 2017. SM50 Sensor Module. <https://goo.gl/FVq4DF>. (2017).
- [2] ams AG. 2017. CCS811 VOC sensor (datasheet). <https://goo.gl/dm08lg>. (2017).
- [3] Matthias Budde, Rayan El Masri, Till Riedel, and Michael Beigl. 2013. Enabling low-cost particulate matter measurement for participatory sensing scenarios. In *Proc. ACM MUM*. 19:1–19:10.
- [4] Saverio De Vito, Marco Piga, Luca Martinotto, and Girolamo Di Francia. 2009. CO, NO<sub>2</sub> and NO<sub>x</sub> urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators B: Chemical* 143, 1 (2009), 182–191.
- [5] Prabal Dutta, Paul M. Aoki, Neil Kumar, Alan Mainwaring, Chris Myers, Wesley Willett, and Allison Woodruff. 2009. Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors. In *Proc. ACM SenSys*. 349–350.
- [6] Jonny Farringdon, Andrew J. Moore, Nancy Tilbury, James Church, and Pieter D. Biemond. 1999. Wearable sensor badge and sensor jacket for context awareness. In *Digest of Papers. Third International Symposium on Wearable Computers*.
- [7] Jill D. Fenske and Suzanne E. Paulson. 1999. Human Breath Emissions of VOCs. *Journal of the Air & Waste Management Association* 49, 5 (1999), 594–598.
- [8] Barbara J. Finlayson-Pitts and James N. Pitts Jr. 1999. *Chemistry of the upper and lower atmosphere: theory, experiments, and applications*.
- [9] Michelle Gallagher, Charles J. Wysocki, James J. Leyden, Andrew Spielman, Xi Sun, and George Preti. 2008. Analyses of volatile organic compounds from human skin. *British Journal of Dermatology* 159, 4 (2008), 780–791.
- [10] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input Using Wrist Gestures. In *Proc. ACM UIST*. 861–872.
- [11] Google. 2017. TensorFlow Mobile. <https://goo.gl/zkZa0B>. (2017).
- [12] Huban A. Gowadia and Gary S. Settles. 2001. The natural sampling of airborne trace signals from explosives concealed upon the human body. *Journal of Forensic Science* 46, 6 (2001), 1324–1331.
- [13] David Hasenfratz, Olga Saukh, Silvan Sturzenegger, and Lothar Thiele. 2012. Participatory air pollution monitoring using smartphones. In *2nd International Workshop on Mobile Sensing*.
- [14] David Hasenfratz, Olga Saukh, and Lothar Thiele. 2012. On-the-fly calibration of low-cost gas sensors. In *Proc. ACM EWSN*. 228–244.
- [15] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, and Lothar Thiele. 2014. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *Proc. IEEE PerCom*. 69–77.
- [16] Simone Herberger, Martin Herold, Heiko Ulmer, Andrea Burdack-Freitag, and Florian Mayer. 2010. Detection of human effluents by a MOS gas sensor in correlation to VOC quantification by GC/MS. *Building and Environment* 45, 11 (2010), 2430 – 2439.
- [17] Jeff Howe. 2006. The rise of crowdsourcing. *Wired Magazine* 14, 6 (2006), 1–4.
- [18] Yifei Jiang, Kun Li, Lei Tian, Ricardo Piedrahita, Xiang Yun, Omkar Mansata, Qin Lv, Robert P Dick, Michael Hannigan, and Li Shang. 2011. MAQS: a personalized mobile sensing system for indoor air quality monitoring. In *Proc. ACM UbiComp*. 271–280.
- [19] Andy P Jones. 1999. Indoor air quality and health. *Atmospheric Environment* 33, 28 (1999), 4535–4564.
- [20] Marc Kamionka, Philippe Breuil, and Christophe Pijolat. 2006. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. *Sensors and Actuators B: Chemical* 118, 1 (2006), 323–327.
- [21] Jung-Yoon Kim, Chao-Hsien Chu, and Sang-Moon Shin. 2014. ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring. *IEEE Sensors Journal* 14, 12 (2014), 4230–4244.
- [22] Silicon Labs. 2017. Thunderboard Sense Kit. <https://goo.gl/oqISyw>. (2017).
- [23] Morton Lippmann. 1989. Health effects of ozone: a critical review. *Journal of Air & Waste Management Association* 39, 5 (1989), 672–695.
- [24] Shengzhong Liu, Zhenzhe Zheng, Fan Wu, Shaojie Tang, and Guihai Chen. 2017. Context-aware data quality estimation in mobile crowdsensing. In *Proc. IEEE INFOCOM*.
- [25] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele. 2016. Pre-Deployment Testing, Augmentation and Calibration of Cross-Sensitive Sensors. In *Proc. ACM EWSN*. 169–180.
- [26] Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele. 2017. SCAN: Multi-Hop Calibration for Mobile Sensor Arrays. *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*. 1, 2 (2017), 19:1–19:21.

- [27] Christian Monn. 2001. Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmospheric Environment* 35, 1 (2001), 1–32.
- [28] Makoto Nakayoshi, Manabu Kanda, Rui Shi, and Richard de Dear. 2015. Outdoor thermal physiology along human pathways: a study using a wearable measurement system. *International Journal of Biometeorology* 59, 5 (2015), 503–515.
- [29] European Research Area Network. 2017. CONVERGENCE: Frictionless Energy Efficient Convergent Wearables for Healthcare and Lifestyle Applications. <https://goo.gl/mGRbRZ>. (2017).
- [30] Evangelos Niforatos, Athanasios Vourvopoulos, and Marc Langheinrich. 2015. Weather with You: Evaluating Report Reliability in Weather Crowdsourcing. In *Proc. ACM MUM*. 152–162.
- [31] Evangelos Niforatos, Athanasios Vourvopoulos, and Marc Langheinrich. 2017. Understanding the potential of human-machine crowdsourcing for weather data. *International Journal of Human-Computer Studies* 102 (2017), 54–68.
- [32] Nima Nikzad, Nakul Verma, Celal Ziftci, Elizabeth Bales, Nichole Quick, Piero Zappi, Kevin Patrick, Sanjoy Dasgupta, Ingolf Krueger, Tajana Šimunić Rosing, and William G. Griswold. 2012. CitiSense: Improving Geospatial Environmental Assessment of Air Quality Using a Wireless Personal Exposure Monitoring System. In *Proc. ACM WH*. 11:1–11:8.
- [33] Lucy Oglesby, Nino Künzli, Martin Rössli, Charlotte Braun-Fahrländer, Patrick Mathys, Willem Stern, Matti Jantunen, and Anu Kousa. 2000. Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution. Results of the European EXPOLIS-EAS Study (Swiss Center Basel). *Journal of the Air & Waste Management Association* 50, 7 (2000), 1251–1261.
- [34] Dinko Oletic and Vedran Bilas. 2015. Design of sensor node for air quality crowdsensing. In *Proc. IEEE SAS*. 1–5.
- [35] Aart Overeem, James C.R. Robinson, Hidde Leijnse, Gert-Jan Steeneveld, Berthold K.P. Horn, and Remko Uijlenhoet. 2013. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters* 40, 15 (2013), 4081–4085.
- [36] Ricardo Piedrahita, Yun Xiang, Nick Masson, John Ortega, Ashley Collier, Yifei Jiang, Kun Li, Robert P. Dick, Qin Lv, Micahel Hannigan, and others. 2014. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques* 7, 10 (2014), 3325.
- [37] Valentin Radu, Panagiota Katsikouli, Rik Sarkar, and Mahesh K. Marina. 2014. A semi-supervised learning approach for robust indoor-outdoor detection with smartphones. In *Proc. ACM SenSys*. 280–294.
- [38] Aakash C. Rai, Chao-Hsin Lin, and Qingyan Chen. 2014. Numerical modeling of volatile organic compound emissions from ozone reactions with human-worn clothing in an aircraft cabin. *HVAC&R Research* 20, 8 (2014), 922–931.
- [39] Rajib Kumar Rana, Chun Tung Chou, Salil S. Kanhere, Nirupama Bulusu, and Wen Hu. 2010. Ear-phone: An End-to-end Participatory Urban Noise Mapping System. In *Proc. ACM IPSN*. 105–116.
- [40] Olga Saukh, David Hasenfratz, and Lothar Thiele. 2015. Reducing multi-hop calibration errors in large-scale mobile sensor networks. In *Proc. ACM IPSN*. 274–285.
- [41] Gerhard Schmitt and Dirk Donath. 2017. ESUM: analysing trade-offs between the energy and social performance of urban morphologies. <http://esum.arch.ethz.ch/about>. (2017).
- [42] Amphenol Advanced Sensors. 2017. Telair T6713 Series CO<sub>2</sub> Module. <https://goo.gl/pSUhu7>. (2017).
- [43] SGX Sensortech. 2014. MiCS-OZ-47 ozone sensor (datasheet). <http://goo.gl/C49tcw>. (2014).
- [44] Olli Seppänen, William Fisk, and Mark Mendell. 1999. Association of ventilation rates and CO<sub>2</sub> concentrations with health and other responses in commercial and institutional buildings. *Indoor Air* 9, 4 (1999), 226–252.
- [45] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical* 215 (2015), 249–257.
- [46] Ewout W. Steyerberg, Andrew J. Vickers, Nancy R. Cook, Thomas Gerds, Mithat Gonen, Nancy Obuchowski, Michael J. Pencina, and Michael W. Kattan. 2010. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)* 21, 1 (2010), 128.
- [47] Philippe Thunis, Anna Pederzoli, and Denise Pernigotti. 2012. Performance criteria to evaluate air quality modeling applications. *Atmospheric Environment* 59, Supplement C (2012), 476 – 482.
- [48] Rundong Tian, Christine Dierk, Christopher Myers, and Eric Paulos. 2016. MyPart: Personal, Portable, Accurate, Airborne Particle Counting. In *Proc. ACM CHI*. 1338–1348.
- [49] PEAN union. 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. (2008).
- [50] Daniel Vallero. 2014. *Fundamentals of air pollution*.
- [51] Chengxiang Wang, Longwei Yin, Luyuan Zhang, Dong Xiang, and Rui Gao. 2010. Metal oxide gas sensors: sensitivity and influencing factors. *Sensors* 10, 3 (2010), 2088–2106.
- [52] Pawel Wargocki, David P. Wyon, Jan Sundell, Geo Clausen, and P. Ole Fanger. 2000. The Effects of Outdoor Air Supply Rate in an Office on Perceived Air Quality, Sick Building Syndrome (SBS) Symptoms and Productivity. *Indoor Air* 10, 4 (2000), 222–236.
- [53] Charles J. Weschler. 2016. Roles of the human occupant in indoor chemistry. *Indoor air* 26, 1 (2016), 6–24.

- [54] Armin Wisthaler and Charles J. Weschler. 2010. Reactions of ozone with human skin lipids: sources of carbonyls, dicarbonyls, and hydroxycarbonyls in indoor air. *Proceedings of the National Academy of Sciences* 107, 15 (2010), 6568–6575.
- [55] World-Health-Organization. 2016. WHO releases country estimates on air pollution exposure and health impact. <https://goo.gl/G4uqFE>. (2016).
- [56] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: when urban air quality inference meets big data. In *Proc. ACM KDD*. 1436–1444.
- [57] Pengfei Zhou, Yuanqing Zheng, Zhenjiang Li, Mo Li, and Guobin Shen. 2012. IODetector: A Generic Service for Indoor Outdoor Detection. In *Proc. ACM SenSys*. 113–126.
- [58] Zhi-Hua Zhou and Ming Li. 2007. Semisupervised regression with cotraining-style algorithms. *IEEE Transactions on Knowledge and Data Engineering* 19, 11 (2007), 1479–1493.
- [59] Xiaojin Zhu and Andrew B Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3, 1 (2009), 1–130.
- [60] Yan Zhuang, Feng Lin, Eun-Hye Yoo, and Wenyao Xu. 2015. Aairsense: A portable context-sensing device for personal air quality monitoring. In *Proc. ACM MobileHealth*. 17–22.

Received May 2017; revised November 2017; accepted January 2018