TOWARDS A STANDARDIZED PROCESS FOR VIRTUAL SENSORS DEVELOPMENT FOR INDOOR AIR QUALITY MONITORING IN DEMOCRATIZED MANUFACTURING

Jorge Merino1, Daiki Ikeuchi2, Alena Puchkova1, Luning Li1,
Sebastian Pattinson2, Stylianos Karatzas1, Ajith Parlikad1
1 Distributed Information and Automation Laboratory, Department of Engineering, University of Cambridge, Institute for Manufacturing, Alan Reece Building, 17 Babbage Road, CB3 0FS, Cambridge, United Kingdom
2 Computer Aided Manufacturing Laboratory, Department of Engineering, University of Cambridge, Institute for Manufacturing, Alan Reece Building, 17 Babbage Road, CB3 0FS, Cambridge, United Kingdom
*jm2210@cam.ac.uk, di261@cam.ac.uk, ap823@cam.ac.uk, ll669@cam.ac.uk,
swp29@eng.cam.ac.uk, sk2095@cam.ac.uk, aknp2@cam.ac.uk

Abstract
Increased accessibility to additive manufacturing technology facilitates democratization of manufacturing, bringing it to habitable environments. The operation of additive manufacturing can be hazardous to human health mid-long term. Virtual sensing extends the capabilities of hardware sensors enabling affordable monitoring to ensure safe operation in democratized manufacturing environments. However, the development process has not yet been standardized for informally trained personnel to facilitate the adoption of virtual sensors. This paper presents a case study analysis to propose a standardized process for the data collection and development of virtual sensors for indoor air quality monitoring in democratized manufacturing environments.

Introduction
Democratization of manufacturing is driven by personal digital fabrication (Mota, 2011; Anderson, 2012; Gershenfeld, 2012), which enables informally-trained consumers to meet their production and customization needs (Browder et al., 2019). Increased access to fabrication technology like Additive Manufacturing (AM) or 3D printing (3DP) brings digital fabrication to ad-hoc manufacturing spaces at home, university, and start-ups that are not designed to manage the harmful emissions for human health (Ford & Despesse, 2016). The widely adopted Material Extrusion (ME) 3DP technology typically utilizes thermoplastic or thermoset polymers as feedstock materials (Ligon et al., 2017), releasing Volatile Organic Compounds (VOC) and Particulate Matter (PM) during their operation. If inhaled, PM (PM2.5 or smaller is smaller than 2 μm) can reach the alveolar region of the lungs and be transported by the bloodstream to other organs (Pilou et al., 2015). (Afshar-Mohajer et al., 2015; Azimi et al., 2017) proved that VOC and PM in indoor environments for AM can greatly exceed the safety thresholds (World Health Organization, 2016).

Indoor environments where AM technology is deployed must be controlled to limit these health hazards since informally trained personnel are neither aware of the health hazards (Berger et al., 2023; Schieweck et al., 2018). Many authors have addressed the characterization of emissions in terms of VOCs and PM from 3D printing processes under controlled laboratory environments (Afshar-Mohajer et al., 2015; Bravi et al., 2019; Kim et al., 2018). Many authors have addressed the characterization of pollutants mostly focus on single sampling of emissions using high-end equipment. Nevertheless, these solutions are not transferable to the built environment (e.g., in offices, classrooms, and makerspaces) due to the reduced number and cost of sampling equipment. In contrast, low-cost sensors are accessible and affordable solutions to monitor the emissions in additive manufacturing and larger environments. The caveat of using low-cost sensors is the lack of accuracy of the available technology to detect emissions. Plus, makers are not willing to add the extra cost of sensors due to their lack of awareness. Some authors try to make low-cost monitoring feasible by integrating it with signal processing and analysis (Stefaniak et al., 2021; Vogt et al., 2021; Martínez-Comesaña et al., 2022; Tagliahue et al., 2021). These techniques can be automated via virtual sensors to increase awareness on IAQ for 3D printing users.

Virtual sensors are online instruments of measurement that supplement other monitoring entities (Lin et al., 2007; Montague et al., 1992; Martin et al., 2021). Virtual sensors have been used to estimate NO, NO2, NOx, SO2, O3, CO, Cd, PM2.5, and PM10 concentration and virtually recreate the pollution in the built environment, in entire neighborhoods (Benabbas et al., 2019; Degoaeu et al., 2019) and undergrounds (Loy-Benitez et al., 2020; Loy-Benitez et al., 2022). Applications of virtual sensing to measure indoor air quality (IAQ) in the literature are still limited to controlled environments (Leidinger et al., 2014; Zaidan et al., 2020), but has proven to be a feasible option for monitoring comfort indoors (Kowli et al., 2023). Virtual sensors can also be used to monitor targets that are out of reach (i.e., it is not possible to install a hardware sensor), and to reduce the number of sensors required making the final solution more affordable. Given the limited literature in the context of democratized manufacturing, there is still no standard process to guide the development of virtual sensors for IAQ.

This paper aims to kick-off the design of a standard process for the development of virtual sensors for IAQ monitoring in ad-hoc democratized manufacturing spaces. The process learns from the ad-hoc solutions in the literature and the analysis of a case study that
implemented a virtual sensor model in a 3D printing environment.

**Context and case study analysis**
The key elements of a case study design are (Priya, 2021):
(a) Purpose of study: To create a data collection protocol and a training and deployment process for virtual sensors to monitor IAQ in a democratized manufacturing environment.
(b) Type of case study research: exploratory case study.
(c) Research question: How to develop and utilize virtual sensors for IAQ monitoring?
The context of the case study is a democratized additive manufacturing environment in a commercial office building. Herein, the study focuses on a cabinet hosting four printers Creality CR20-PRO Fused Deposition Modelling (FDM), an ME technique. FDM 3D printing principle consists in building three-dimensional solid objects from their digital models by selectively accumulating liquified thermoplastic material (e.g., PLA, ABS) layer-by-layer (Anon, 2024; ISO/ASTM, 2013) into a bed where the material cools and cements, shaping a geometry. The cabinet is enclosed for the sake of safety in terms of limiting the emissions released into the rest of the shared working environment; yet this space represents a practical democratized manufacturing space in terms of size, purpose, and the level of available resources. The dimensions of the cabinet are 200x90x100 cm (see Figure 1). It is composed of a desk and eight methacrylate panels for two side panels, four access windows, two sliding windows in the back and the ceiling, plus a single extraction point in the center of the ceiling panel. Extraction works at the same speed but can be turned on and off. The goal of monitoring in this space is to raise awareness on health for the users in terms of utilization of the space, and the cabinet, and pollution escaping the cabinet during operation of the 3D printers.

**Virtual sensor data collection, training, and deployment process**
This section explains the process for collecting IAQ data and training and deploying virtual sensors in the democratized additive manufacturing environment. This process is addressed to occupants operating democratized additive manufacturing environments and data practitioners interested in monitoring air quality indoors.

![Figure 2. Process for data collection and development of Virtual sensors for IAQ monitoring.](image)

Figure 2 summarizes these steps in the process, which are the result of the analysis of literature on IoT deployments (Bushir et al., 2022; Bibri, 2018), prior knowledge in other areas (Merino et al., 2022), and adaptations from the application of each step to the domain of IAQ.

**Low-cost sensor deployment**
An effective sensor deployment is launched by appropriate selection of sensors. Sensor selection must be informed by literature in the domain as well as a market review. For instance, in the context of 3DP, (Afshar-Mohajer et al., 2015; Bravi et al., 2019; Kim et al., 2015; Mendes et al., 2017; Pernetti et al., 2023) used chamber setups to characterize the emissions of popular 3DP filament materials like PLA, ABS, HIPS, and PC, resulting in particulate matter (PM) and diverse species of volatile organic compounds (VOCs).

Sensors shall be calibrated before deployment. Although manufacturers claim that sensors have been calibrated in laboratory environments at source, they need to be calibrated in the field since fabrication variance is not avoidable yet in low-cost sensing. The calibration methods for air quality monitoring vary (Chowdhury et al., 2023; Zaidan et al., 2020; Zarrar & Dyo, 2023; Tariq et al., 2020; Zimmerman et al., 2018), but share the notion of comparing against a source of truth. Often these sources of truth are high-quality sampling equipment.
This equipment comes at a high price tag both for purchasing and for single-time calibration from the manufacturer, which may be prohibitive for some makers. Furthermore, concept drift (De Vito et al., 2024) appears when operating conditions change from those encountered during training set recording, the derived calibration accuracy drops. Thus, calibration should be considered over longer-time of monitoring to account for seasonal differences and sensor drifts. Alternatively, some authors targeting scalable calibration have proposed median calibration, which is obtained through using datasets considering the median response of several sensor units to identical field recorded conditions (De Vito et al., 2024).

Finally, the position of the low-cost sensors in the space must be designated for deployment. Sensor positioning in air quality monitoring is informed by the identification of sources of pollution and the expected usage of data. The final positions and configuration of sensors must be logged into a repository for traceability and maintenance of the deployment, since hardware sensors can suffer network drops and reconfigurations, power problems, or even complete failure that prompts replacement. For deployments with limited number of sensors, can make them rotate around the intended positions, then later substituted for their virtual counterparts.

**Data Collection: Characterization of pollutants in a practical space**

The purpose of the data collection is to characterize how the pollutants are dispersed in a practical democratized additive manufacturing environment.

Variables are selected to enable virtual sensing training and operation. Generally, nearby locations measurements are used to estimate PM and VOC, but some authors have used humidity and temperature to improve the accuracy of CO2 estimations (Tagliafue et al., 2021; Zaidan et al., 2020), CO and CO2 and operational parameters of the 3D printers for estimating PM and VOC (Stefaniak et al., 2021).

Experiments must be contextualized in the process of the indoor environment that is causing the emissions. The goals of data collection and literature can support the design of experiments. In the case of semi-controlled or uncontrolled environments, experiments conducted shall cover data collection of a range of practical situations. Experiments should have the same starting conditions and they are repeated for validation of data collection under the same conditions. Experiments can also be repeated for identification of seasonality.

Data analyzed to test the validity and thoroughness of data collection and to identify potential missing conditions for new experiments. Experiments ought to include at least baseline identification (i.e., how is the signal for each variable when nothing is happening) and the range of practical situations devised during experiment design.

**Virtual Sensing design**

Virtual sensors are used as a backup for potential failures of low-cost sensors, to completely replace them, and to provide measurements in locations hardware sensors cannot reach (e.g., temperature inside a boiler). In the case study, the virtual sensor was designed as a backup for PM1.0 of the low-cost sensor in the same location and to replace it for use in a different space.

Diverse regression and machine learning models have been used to estimate PM and VOC: in terms of indoor air quality, (Stefaniak et al., 2021) used linear regression to estimate VOC and PM concentration in a large-format yet controlled AM lab. (Dogeau et al., 2019) trained Neural Network virtual sensors on data from fixed air quality stations located in urban areas to estimate the concentration of PM2.5 and PM10 among other pollutants to virtually recreate the pollution in entire neighborhoods. (Leidinger et al., 2014) implemented a virtual sensor based on linear discriminant analysis (LDA) for selective detection of hazardous VOCs using a gas sensor array. Literature can influence model selection, but it is ultimately the decision of the data practitioner which model to select. Comparison between several models is recommended. It is important to understand that models trained in a given space may not be transferable to other spaces unless the geometry and operations are near to identical (e.g., another cabinet with the same setup). Models in the literature should be implemented within the context of the case study to test the validity and enable comparison under desired conditions.

Along with model selection, comes the features to be used to train the virtual sensors. Data analysis should undercover correlations between variables that can support estimations in the models. Literature can also inform known relationships between variables, but they should be confirmed through data analysis.

In most cases, raw data from hardware sensors needs some form of preparation before they are suitable for model training. Typical methods for data preparation are filtering, synchronization, and transformations (e.g., principal component analysis, logarithmic conversion, or derivatives of the original signal). Outliers due to sensor temporary malfunction and missing data due to power or network losses should be identified and removed, and only relevant controlled data from the experiments should be used for training.

With the advent of Machine Learning open-source frameworks, model training is becoming easier than ever. (Nguyen et al., 2019; Wu et al., 2022) compare popular frameworks in terms of capabilities and training and testing performance. The main goal of training is to obtain acceptable accuracy without incurring into overfitting models. This becomes an iterative process of trial and error, and continuous improvement in most cases. In some cases, this iterative process requires the practitioner to apply different pre-processing methods, selecting different features, or conducting further data analysis.
Virtual Sensing deployment

Having a trained model that can estimate the pollutants in a given position is not enough for using a virtual sensor in a practical case. Virtual sensors for IAQ monitoring need to be understood as standalone systems and behave similarly to low-cost sensors. A system architecture is necessary to deploy virtual sensors in a scalable manner (Merino et al., 2023; Martin et al., 2021; Leidinger et al., 2014; Zaidan et al., 2020).

Any type of virtual sensor needs evaluation (Goodwin, 2000), but more importantly when it comes to IAQ monitoring since they are aimed to monitor pollutants that can affect human health. Virtual sensors need long-term calibration (Koo & Yoon, 2022) since their accuracy can decrease because of various uncertainties in the system operation and virtual sensor model.

Case study application and Discussion

This section describes the application of the data collection, training, and deployment process for virtual sensors. Results of implementation are discussed including the identification of the challenges faced and adaptations made to the process.

Sensor deployment implementation

The literature on air quality in additive manufacturing environments guided the selection of sensors. The materials used in the 3D printers are PLA and ABS, which emit PM and diverse species of VOC (Steinele, 2016). The virtual sensor designed focuses on PM. Low-cost sensors available in the market to monitor these two variables are based on the light-scattering principle for PM. Sensors from two manufacturers are used: M1 can measure PM1.0, PM2.5, and PM10, and M2 can monitor PM1.0, PM2.5, PM10, tVOC, Temperature, Humidity, CO2, and Ozone.

Sensors were calibrated using the median approach. While other authors installed sensors in their final position to collect data for median calibration, sensors were collocated in stacks in the same center of the desk inside of the cabinet and shuffled once during the calibration data collection. This deviation from the literature moved the calibration step before the sensor positioning in the process. Median was calculated across all sensors and selected as a general calibration law. A data preprocessing module was added to the data pipeline to locally adjust the differences between each sensor signal and the median. Additionally, a sensor drift analyzer was set up to identify and correct potential sensor drifts due to aging. Calibration was performed per manufacturer.

The positions of the sensors are determined by the positions of the 3D printers in the cabinet. There are four 3D positions for the 3D printers evenly distributed along the Y axis (200 cm long) in the cabinet, centered at 25 cm, 75 cm, 125 cm, and 175 cm. The purpose of devised virtual sensors is to estimate distribution of pollutants within the cabinet, therefore, the sensors (four M1, and fifteen M2) were evenly distributed in the cabinet.

Data collection implementation

Temperature, Humidity, and PM1.0 from all sensor locations are considered. Temperature and humidity have an impact and Particulate matter sensors, respectively and can help to identify seasonal bias.

The following experiment variables were selected: extraction of the cabinet (i.e., can be on/off, when it is on, it always extracts air at the same speed); windows opened (i.e., whether it is planned to open any window during the experiment, monitored by open closed sensors); printer positions in operation (i.e., takes the form of an array of Booleans, e.g., [1, 1, 0, 0] means that only printers in positions 1 and 2 are printing); material (i.e., type of material used throughout the printing, either PLA or ABS from the same manufacturer and model in all the experiments); and nozzle temperature. The experiments consisted of permutations of the variables. For example, experiment 1 included extraction off, windows closed, and 3D printer in position 1 printing a geometry using ABS for one hour. Before each experiment, the air inside the cabinet is exchanged for at least one hour. Experiments with the same conditions (i.e., variable setups) were repeated at least three times.

The baselines of PM1.0 in all sensor locations were identified over a period of 7 hours under non-printing conditions by repeating the experiment with windows open and closed, extraction on and off. For instance, Figure 3 depicts two baselines identified for PM1.0 in the target location (i.e., the sliding window used to access the printed geometries): Extraction Off and door closed (400-420 ppm), and extraction on/windows opened (380-400 ppm). This enabled a baseline correction reference for each sensor parameter in data.

Among the permutation of experiment variables under printing conditions, Figure 4 depicts the experiments for position 1 printer printing with PLA over a period of 1h:30’. Some conclusions of this plot are that extraction is confirmed to help with the clearance of the cabinet, both in concentration of PM and time for the concentration to decrease back to baseline after the end of the experiments.
Having two different sensor manufacturers required repeating the analysis for repeated parameters. Sensors M1 were more reactive to changes in PM concentration than M2. Cross-manufacturer comparisons are required. Data collection is an iterative and cyclical process even with planning. Furthermore, some data collection impacted the selection of more sensors of Manufacturer M1 as well as the position (together with sensors M2).

**Virtual sensing design implementation**

Several models were selected both linear (e.g., lasso, elastic networks, Bayesian) and machine learning (e.g., boosted trees). Linear regression models have fair accuracy when supported by online measurements from hardware sensors (Stefaniak et al., 2021), whereas boosted trees have been used to estimate PM in (Koo & Yoon, 2022). A simple case is to remove the time variable of the series and estimate individual values based on the measurements of the reference low-cost hardware sensor in the same space, for which all linear models performed (Merino et al., 2023). This model lacks the capability of identifying trends since it is not based on time-series estimation.

Features selected for PM estimation are measurements of the same PM size from surrounding hardware sensors. For the simple case, the correlation between each of the PM variables (i.e., PM1.0, PM2.5, and PM10) is enough to train linear models when estimating single values. Thus, to estimate PM1.0 in sensor a target position, PM1.0 readings from all positions but the target are selected. This decision is in line with the purpose of the virtual sensor, to serve as a backup for and potentially replace the target low-cost sensor location (i.e., by the sliding door).

3D printer logs and low-cost hardware sensors required timestamp synchronization to compensate for the skew derived from internal clock drifts and different time server setups of the sensor and 3D printer platforms. Additionally, outliers identified (e.g., sudden jumps in raw signal or subsequent out-of-range measurements due to malfunction of PM and VOC sensors) are removed if the value is beyond two standard deviations. Uncontrolled data is discarded for the purposes of the virtual sensor. Finally, a Savitzky-Golay filter (Savitzky & Golay, 1964) was applied to smooth the curves. For the simplest case of the single-value estimations based on surrounding sensors, the time dimension is removed. Since the signals are previously synchronized that results into a set of arrays with values for each sensor position at a given time.

Models trained take the PM filtered and untyped arrays and having partitioned them into 80% training and 20% testing segmented sub-sets. Linear models R² scores reached 0.826 for Lasso, 0.826 for Elastic Networks, and 0.826 for Bayesian ridge for the simple approach. Non-linear approach using temperature and/or humidity as supporting estimators for both PM could not reach a significant R² score to be considered good estimators in a practical case, potentially due to the transformation into single-value estimation rather than time-series estimation, which enable trend identification and more accurate regression. Trained models can estimate the pollutant concentration in a specific point where the original low-cost sensor was positioned, therefore, they can be used to virtualize.

**Virtual sensing deployment**

The architecture published in (Merino et al., 2023) was designed to streamline virtual sensors training and operation. The architecture modularizes the process to facilitate data collection, preprocessing, training, model storage and loading, virtual sensor execution, and recalibration. In the simple approach for estimating PM based on surrounding sensors signals, the evaluation consisted in comparing the signal from the virtual sensor against the original sensor used to train the model after a period of dual operation (i.e., both the virtual sensor and the low-cost hardware sensor measuring PM values for a period of two weeks before replacing the hardware sensor).

The need for recalibration of virtual sensors can be identified by deploying a hardware sensor again in the original position. In a practical democratized additive manufacturing environment, or any other indoor building operation, hardware sensor may not always be available for recalibration, therefore, other approaches like (Koo & Yoon, 2022) may be more suitable.

**Conclusions**

Indoor air quality monitoring in democratized additive manufacturing environments is still not sufficiently explored. Virtual sensing approaches have demonstrated good performance for outdoors air quality and indoor comfort and pose a good opportunity to improve the awareness and control of hazardous pollutants. Beyond increasing awareness of emissions indoors at an affordable price, virtual sensors have the potential of informing ventilation systems operation to ensure operators safety while minimising building energy management. Virtual sensors can help understanding spatiotemporal distribution of emissions and emission flow into the rest of the built environment (e.g., adjacent offices, classrooms, or even bedrooms at home). Furthermore, IAQ monitoring can have the potential to inform on materials and design of indoor spaces as well as maintenance during the operation phase of the built environment (Abdalla & Peng, 2021).

This paper proposes a standard process for data collection and development of indoor air quality using virtual sensors in a practical democratized additive manufacturing environment, resulting of the analysis of the literature and the adaptations after the implementation in a practical democratized additive manufacturing environment. A simple example of a virtual sensor is explained throughout the implementation. The aim of the process is to facilitate the entry barrier to this technology to ad-hoc democratized manufacturing spaces, answering the research question. Additionally, the process should
also address a second research question in the next iteration, on how to overcome lack of trust in low-cost monitoring and consequently in virtual sensing. Evaluation and calibration approaches should be investigated further to answer this question.

This process needs to be implemented in a set of case studies in similar democratized manufacturing spaces to test its validity and find adjustments. Additionally, a replication of this study in other democratized manufacturing environments would be beneficial for the confirmation of findings and to discover the casuistic across different setups and additive manufacturing technologies. Another limitation of this study is within the fact that indoor environments suffer from pollution from non-characterizable sources like outdoors pollution from construction sites or traffic. Case studies identified to complement this work in this direction and to ensure full transferability to the built environment context include domains like accommodation, offices, hospitals, universities, and public transport stations.

Acknowledgements

This work was supported by the UK’s Department for Digital, Culture, Media and Sport through the project entitled “Digital Twins Enabled IAQ Management for Healthy Living-TwinAIR”. TwinAIR is funded by UK Research and Innovation (UKRI) through the Horizon Europe Guarantee Scheme for European Union’s Horizon Europe Research and Innovation Programme, grant number 101057779, HORIZON-HLTH-2021-ENVHLTH-02-02.

References


Goodwin, G.C. (2000) Predicting the performance of soft
sensors as a route to low cost automation. Annual Reviews in Control, 24, pp.55–66.


