Technische Universität München
TUM School of Engineering and Design

Machine Learning-Based Virtual Sensing of Indoor Air Pollutants: Enabling Demand-Controlled Ventilation for Improved Indoor Air Quality and Energy Efficiency

Martin Gabriel

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Klaus Peter Sedlbauer

Prüfende der Dissertation:

1. Prof. Dipl.-Ing. Thomas Auer
2. Prof. Dr.-Ing. Frank Petzold
3. Prof. Dr. Thomas Wortmann

Die Dissertation wurde am 09.01.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 11.06.2024 angenommen.

Danksagung

Die vorliegende Arbeit wurde während meiner Zeit als Doktorand bei der ENIANO GmbH sowie am Lehrstuhl für Gebäudetechnologie und klimagerechtes Bauen der School of Engineering and Design an der Technischen Universität München angefertigt.
Mein aufrichtiger Dank gilt Prof. Thomas Auer, der die Betreuung meiner Dissertation übernommen hat. Der fachliche und persönliche Austausch war außerordentlich bereichernd und hat maßgeblich zum Erfolg meiner Forschung beigetragen. Ebenso danke ich Prof. Dr. Frank Petzold für seine fachliche Unterstützung.
Weiterhin bedanke ich mich herzlich bei Tobias Eder, den Geschäftsführer der ENIANO GmbH, für die Ermöglichung meines Promotionsvorhabens sowie die stets anregenden Diskussionen über die Ideen und Methoden meiner Arbeit. Ein herzliches Dankeschön geht an Dr. Jasmin Gärtner, die als Mentorin meiner Dissertation fungierte. Ihre Unterstützung und der stetige Austausch waren äußerst wertvoll. Mein Dank gilt ebenso Klaus Greiling für das Ermöglichen der Messungen und den interessanten Einblick in die Praxis. Vielen Dank auch an meine KollegInnen bei ENIANO und der Technischen Universität München für die spannenden und inspirierenden Gespräche.
Ein besonderer Dank geht an meine Familie, die mich auf meinem gesamten Bildungsweg stets unterstützt hat. Ohne diese fortwährende Unterstützung wäre diese Arbeit nicht möglich gewesen. Abschließend möchte ich Kathrin für ihre Geduld, Ermutigung, Unterstützung und den Ausgleich in allen Phasen dieser Dissertation danken.

Abstract

Monitoring individual exposure to indoor air pollutants is crucial for human health and well-being, as well as for energy-efficient HVAC operation. However, the high spatiotemporal variations of indoor air pollutants make ubiquitous sensing essential. Currently, physical sensors are costly to maintain, making this approach infeasible. Therefore, this study aims to investigate the feasibility of virtually sensing indoor air pollutants, such as particulate matter, volatile organic compounds (VOCs), and CO2, using machine learning methods.

To achieve this, several years of accumulated measurement data from multiple zones and typologies of non-residential buildings were collected using customized measurement nodes. This data was used to evaluate pollutant dynamics and concentration levels in the examined rooms, as well as to train machine learning models. These models predict indoor air pollutant concentrations based on Building Management System (BMS) data, including temperature, humidity, illumination, noise, motion, and window state, as well as meteorological and outdoor pollution data. A cross-validation scheme and hyperparameter optimization were employed to determine the best model parameters and evaluate performance using common evaluation metrics such as mean absolute error (MAE) and root mean square error (RMSE).

Various machine learning methods were examined, with a specific focus on long short-term memory (LSTM) networks. Other algorithms, such as multilayer perceptrons (MLPs) and stochastic gradient descents (SGDs), were found to be less effective when supplied with large amounts of data. The transferability of the models to other rooms, buildings, and typologies was evaluated both for unseen environments and using a transfer learning approach. The results indicated that the models were transferable within the same typology. However, for other typologies, transfer resulted in significantly diminished prediction performance, highlighting the need for a broader training dataset to improve generalization.

Two case studies were conducted to assess the applicability of the virtual indoor air pollutant sensor in demand-controlled ventilation. One case study utilized a calibrated simulation model, while the other involved a decentralized ventilation unit in a classroom. The case studies demonstrated a significant potential for energy reduction by deploying virtual indoor air pollutant sensors. Depending on the building's typology and the previously used control method, the results indicated a potential reduction in air treatment and transportation energy consumption of up to 95%. These reductions could be achieved by minimizing oversupply during unoccupied or low-occupied times while ensuring sufficient airflow during occupation, thus maintaining indoor air quality.

Zusammenfassung

Sowohl für die Gesundheit der Gebäudenutzerinnen und -nutzer als auch für den energieeffizienten Betrieb von Lüftungsanlagen ist die Überwachung der Luftschadstoffbelastung in Innenräumen entscheidend. Insbesondere aufgrund der hohen räumlichen und zeitlichen Variation von Luftschadstoffen in Innenräumen ist eine räumliche und zeitliche Abdeckung durch Sensorik unerlässlich. Allerdings machen die Kosten und der Wartungsaufwand von physischen Sensoren dies derzeit in der Baupraxis unpraktikabel. Deshalb untersucht diese Dissertation die Anwendbarkeit von virtueller Sensorik - basierend auf Machine Learning Methoden - für die Erfassung von Luftschadstoffen (Feinstaub, VOC, CO2) als Alternative zu physischen Sensoren. In diesem Rahmen wurden über mehrere Jahre Messdaten aus verschiedenen Räumen, Gebäuden und Gebäudetypologien mit eigens entwickelter Sensorik gesammelt. Die gesammelten Daten wurden zur Beurteilung der Schadstoffdynamik und Konzentrationsniveaus verwendet sowie zum Training der Machine-Learning-Modelle. Die Modelle prognostizieren die Schadstoffkonzentration basierend auf Daten des Gebäudemanagementsystems (z.B. Temperatur, Feuchtigkeit, Beleuchtung, Lärm, Bewegung und Fensterzustand) sowie meteorologischen Daten und der Feinstaubbelastung im Außenraum. Die virtuellen Sensoren wurden mithilfe gängiger Evaluationsmetriken (MAE, RMSE) bewertet.
Verschiedene Machine Learning Methoden wurden in diesem Rahmen untersucht, wobei der Schwerpunkt auf Long-Short-Term-Memory (LSTM) Netzwerken lag. Die Ergebnisse zeigten, dass andere Machine Learning Algorithmen wie Multi Layer Perceptrons (MLP) oder Stochastic Gradient Descent (SGD) bei einem großen Trainingsdatenvolumen schlechtere Ergebnisse lieferten. Die Übertragbarkeit der trainierten Modelle auf andere Räume, Gebäude und Typologien wurde sowohl ohne Anpassung der Modelle als auch mit einem Transfer-Learning Ansatz getestet. Die Ergebnisse zeigten eine gute Übertragbarkeit innerhalb derselben Typologie. In einer anderen Gebäudetypologie nahmen die Vorhersageergebnisse hingegen stark ab und deuten auf die Notwendigkeit eines größeren Trainingsdatensatzes hin, um eine bessere Generalisierbarkeit des Modells zu erreichen.
Die Integration der virtuellen Luftschadstoffsensoren in die Steuerung bedarfsgerechter Lüftung wurde anhand zweier Fallstudien getestet - zum einen in einem kalibrierten Simulationsmodell eines Großraumbüros, zum anderen anhand einer Steuerung eines dezentralen Lüftungsgerätes in einem Klassenzimmer. Die Fallstudien zeigten ein erhebliches Einsparpotenzial von bis zu 95% der Energie der Lüftungsanlage. Die Einsparpotenziale können durch die Minimierung von Überversorgung während geringer oder keiner Belegung erschlossen werden, während gleichzeitig während hoher Belegung für ausreichenden Luftwechsel und somit eine gute Innenluftqualität gesorgt wird.

Contents

List of Figures

List of Tables

Acronyms

BMS	Building Management System.
ETL	Extract Transfer Load.
GPU	Graphics Processing Unit.
HVAC	Heating, Ventilation, and Air Conditioning.
IAP	Indoor Air Pollutants.
IAQ	Indoor Air Quality.
IAQI	Indoor Air Quality Index.
IoT	Internet of Things.
LSTM	Long Short-Term Memory Network.
MAE	Mean Absolute Error.
MLP	Multi-Layer Perceptron.
MOS	Metal Oxide Sensing.
NDIR	Non-Dispersive Infrared.
OPC	Optical Particle Counter.
PM	Particulate Matter.
ppm	Parts Per Million.
RMSE	Root Mean Squared Error.
SVM	Support Vector Machine.
VOC	Volatile Organic Compounds.

Part I

Theory

1 Introduction

1.1 Relevance of Air Quality

Indoor air pollutants come in different sizes, types, and aggregate conditions (liquid, solid, or gaseous) with diverse chemical attributes and intake routes. These pollutants can be categorized into inorganic gases, organic gases, particulates, microbial pollutants, bacteria and viruses [Tham, 2016]. As we spend around 87% of our time indoors [Hasager et al., 2021], controlling indoor air pollutant exposure becomes crucial, as buildings become the primary sources of exposure. This importance has been further emphasized by the increasing energy efficiency of buildings, which results in higher sealing and self-contained indoor environments with a high dependence on HVAC systems [Hasager et al., 2021].

Research has shown that poor indoor air quality can cause various symptoms and health issues, with the severity of these effects depending on factors such as pollutant type, concentration, exposure duration, and individual characteristics like age, health, and constitution [Berglund et al., 1992]. Common symptoms include fatigue, eye irritation, headaches, and difficulty concentrating [Berglund et al., 1992]. Prolonged exposure can lead to respiratory disorders and immune responses, such as asthma, especially among vulnerable groups like children or the elderly [Berglund et al., 1992].

Research on outdoor air pollution has already demonstrated the health risks associated with polluted air. A WHO study found that increased particulate matter concentration was linked to 400,000 premature deaths in European countries [Soares, 2020]. Consequently, many countries are implementing measures to reduce air pollutant levels through concentration and exposure time limits. In addition to enforcing these limits, research is also exploring effective strategies for reducing indoor air pollutant concentrations. These strategies include source control, which aims to identify and minimize sources of pollution, as well as mitigation measures such as filtration to remove pollutants and introducing clean air, and monitoring pollutant concentrations to reduce occupants' exposure time accordingly [Tham, 2016].

The applicability of these measures varies across the building lifecycle. Source control is mainly relevant in the planning and construction phase of buildings, while mitigation and monitoring are relevant during the operational period. Source control in new buildings is progressively ensured by increasing standards in building codes that restrict the use of harmful, pollutant-emitting materials. However, most buildings in the existing building stock were constructed without such regulations in place. Therefore, it is crucial to focus on mitigation and monitoring in the existing building stock, especially in non-residential typologies, defined as buildings not intended for living or accomodation purposes, where occupants usually have limited control over the indoor environment compared to residential buildings. This study explicitly focuses on open office and classroom typologies

as they represent common non-residential typologies where individuals spend significant amounts of time and where the indoor environment significantly impacts occupant comfort, productivity, and health.

1.2 Motivation and Problem Statement

Indoor air pollutant concentrations can be up to 100 times higher than outdoor concentrations, which increases the risk of diseases such as cancer, respiratory disorders, and immune system disorders, resulting in an estimated 2 million premature deaths per year [Smith and Mehta, 2003]. Office buildings in industrialized nations, in particular, pose a significant risk due to frequent occupancy. The World Health Organization (WHO) estimates that 30% of the global stock of office buildings have substandard air quality, negatively impacting the health and well-being of 10-30% of occupants. The significance of indoor air quality has been acknowledged for over a century [Sundell, 2004], but efforts to enhance energy efficiency have unintentionally worsened the issue. Measures such as airtight building envelopes, reduced air exchange, and recirculation systems have led to the accumulation of pollutants and the spread of infectious diseases. However, the lack of effective methods for quantifying indoor air quality has resulted in inadequate consideration in building systems control. Fortunately, the increasing availability of data in buildings and advancements in air pollutant sensor systems offer the potential for precise and widespread measurement of indoor air pollutants throughout a building. Recent studies have also demonstrated significant spatiotemporal variations in indoor air pollutants, reinforcing the need for high-resolution monitoring to accurately assess risks. Unfortunately, the scalability of such measurements is currently limited by the sensitivity and high cost of sensor equipment. Therefore, alternative approaches to physical sensing must be developed in order to achieve scalable indoor air pollutant monitoring.

1.3 Research Objectives and Scope

Virtual sensing, which aims to approximate unmeasured physical quantities in a dynamic system using existing sensor information, offers an alternative to ubiquitous sensor deployment. It typically employs physical modeling techniques (white-box), hybrid methods combining both modeling and data-driven elements (grey-box), or machine learning approaches (black-box) [Heindel et al., 2021]. Virtual sensing has found applications in process control, automotive, avionics, and robotics [Li et al., 2011]. With the increasing digitalization of buildings through the use of Internet of Things (IoT) devices, the volume of data generated from sensors, building systems, and networked devices continues to rise. This availability of high-resolution spatial and temporal data presents an opportunity to adapt virtual sensing in the built environment. Therefore, the objective of this dissertation is to develop a long-term indoor air pollutant dataset for various rooms and building typologies. Using this dataset, the potential of virtual sensing for indoor air pollution estimation will be explored by training machine learning models on the accumulated data. These models' performance will be evaluated across different rooms and typologies

under a wide range of conditions. Additionally, the feasibility of implementing virtual indoor air pollutant sensors in demand-controlled ventilation systems will be examined. The potential of these virtual sensors will be investigated in case studies, assessing their impact on air treatment and transportation energy consumption when deployed in demand-controlled ventilation systems.

1.4 Research Questions and Hypotheses

The literature review will address the following questions:

1. What is the current state of research on indoor air pollution in non-residential buildings, and how does it vary across rooms, seasons, and building types?
2. What methods are available for long-term, high spatiotemporal resolution, on-site measurements of indoor air pollutants? What are their limitations and applications?
3. Are there existing methods for virtual sensing of indoor air pollutants? What implementations for indoor air pollutant monitoring exist?
4. What research has been done on control strategies for demand controlled ventilation and the integration of indoor air pollutant measurements?

Based on the findings from the literature review, this dissertation aims to answer the following research questions:

1. What is the state of indoor air pollution in classroom and open office zones? What are the seasonal variations, correlations, and distribution patterns in these zones, and how do they compare?
2. Can machine learning techniques be used to develop multi-pollutant virtual sensors for predicting indoor air pollutant concentrations? Can these virtual sensors be transferred to different zones and building typologies?
3. Can virtual sensors be integrated into demand-controlled ventilation systems? What are the potential air treatment and transportation energy consumption reductions and associated benefits of using virtual indoor air pollutant sensors in non-residential buildings?

Based on these research questions, the following hypothesis guides this dissertation: *Virtual indoor air pollutant sensors, which utilize machine learning models (LSTM, MLP, SGD), have the potential to accurately forecast indoor air pollutant concentrations ($PM_{2.5}$, CO_2, VOC).*

1.5 Methodological Approach

This dissertation will employ the following methods:

Literature Review A comprehensive review of the literature will examine the current state of research on indoor air pollutants in non-residential buildings, physical sensing versus virtual sensing of these pollutants, and demand-controlled ventilation based on indoor air pollutants. The literature review will identify gaps in the existing research.

Dataset Creation An extensive dataset on indoor air pollutant concentrations and building management system (BMS) data will be compiled using sensing methods identified during the literature review. The dataset will include multiple years of data from various rooms and building typologies.

Dataset Processing and Analysis The dataset will be processed and analyzed using big data techniques. Statistical methods will be used to identify trends, variations, and correlations between indoor air pollution, building typologies, and environmental factors.

Machine Learning Machine learning techniques will be employed to develop virtual sensing models for predicting indoor air pollutants in non-residential buildings. The dataset will be processed and optimized for machine learning models. Model training and hyperparameter optimization will be conducted. The models will be evaluated using standard metrics, and their transferability to unseen conditions will be tested. A comparative analysis of different machine learning models will also be performed. .

Integration in Demand-Controlled Ventilation The potential air treatment and transportation energy savings of integrating the virtual sensing model into a demand-controlled ventilation system will be evaluated through two case studies conducted in real-world environments. These case studies will utilize calibrated simulation models and/or direct control of the ventilation system. Numeric building energy simulation tools will be employed for these simulations, comparing the indoor air pollutant-based control strategy to other prevalent ventilation control strategies.

1.6 Structure of the Dissertation

The dissertation is structured into five main parts: the theory, the methods, the results, the case studies, and the discussion and conclusion. The theory section discusses the fundamentals of building ventilation, indoor environmental quality, and indoor air pollutants. It concludes with a state-of-the-art review of current literature on indoor air pollutants in non-residential buildings, measurement methods, virtual sensing of pollutants, and demand-controlled ventilation. The theory section also identifies the desired contribution of this dissertation to existing research. The methods section begins with the construction of a dataset, including the measurement setup, locations, equipment, and quality control. It then proceeds to the preprocessing of the dataset, the machine learning algorithms used, evaluation methodologies, and testing for model transferability and transfer learning. The results section presents the dataset characteristics, providing a breakdown for individual rooms where measurements were taken. It then presents

the results of the virtual sensing model, including model evaluation, transferability evaluation, transfer learning evaluation, comparison to other machine learning algorithms, and a discussion of the findings. The case studies provide a detailed description of the application of the virtual sensing model in demand-controlled ventilation in two different environments - an open office and a classroom. Each case study includes information on the setup, framework, calibration, control strategies employed, and the results obtained. The discussion and conclusion section summarizes the key findings, interprets the results, compares them to related works, outlines the strengths and limitations of the study, discusses implications for practice and policy, and provides recommendations for future research. This section also recaps the research questions and findings, highlights the contributions of the dissertation to the field, offers final thoughts, and presents a future outlook.

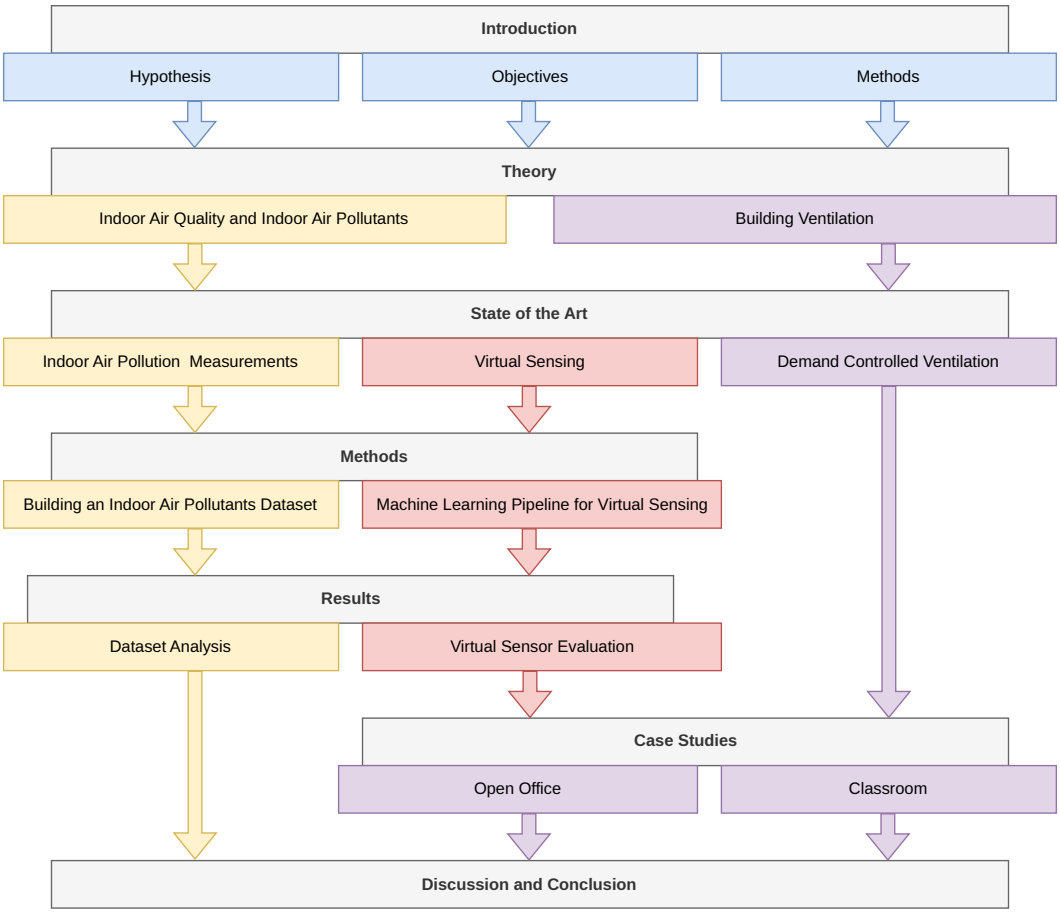


Figure 1.1: Visual representation of the dissertation structure

1.7 Contributions and Expected Outcomes

This dissertation aims to contribute to the literature and expects the following outcomes:

1. Development of a comprehensive, long-term, high spatiotemporal resolution indoor air pollutant dataset that includes multiple rooms and building typologies.
2. Implementation and evaluation of machine learning techniques to develop multi-pollutant virtual sensing models as potential substitutes for physical indoor air pollutant sensors.
3. Demonstration of the transferability and adaptability of virtual sensing models across different rooms and typologies.
4. Improvement of the energy efficiency of demand-controlled ventilation systems through the integration of virtual indoor air pollutant sensors.

The findings of this research can have potential applications within the building industry and significantly contribute to the current state of research. The expected contributions and applications include:

1. Utilization of virtual sensing technology for real-time occupant exposure monitoring, informing stakeholders, and enabling adaptive indoor air pollutant management strategies.
2. Improvement of building operations and energy efficiency by integrating virtual sensors into ventilation control systems, thereby minimizing oversupply.
3. Enhancement of occupant health, well-being, and productivity through informed decision-making and targeted indoor air pollutant interventions in non-residential buildings.

2 Fundamentals

The following chapter presents the fundamentals of building ventilation, indoor air quality within indoor environmental quality, and the most relevant indoor air pollutants, their impacts, and regulated thresholds.

2.1 Building ventilation

The following section examines the fundamentals of building ventilation, looking at the different approaches to building ventilation and the topology, components, and control of mechanical ventilation systems.
By definition ventilation is the technical approach for the controlled movement of air in buildings. The main goal is to supply fresh air and extract old air from the interior by moving the air [Hausladen and Tichelmann, 2012]. Air movement is induced through pressure differences in the interiors, between interior and exterior, and by atmospheric pressure differences from wind and weather patterns [Hausladen and Tichelmann, 2012]. In the building context, pressure differences are either passively induced through physical phenomena or actively created by mechanical systems.

2.1.1 Building ventilation types

Building ventilation approaches can be differentiated into natural ventilation (a), extract (b), and supply and extract systems (c). These approaches are differentiated by the strategies they employ in order to extract old air and supply fresh air.

Natural ventilation (a) Natural ventilation is the direct exchange of inside and outside air without active technical equipment [Hausladen and Tichelmann, 2012]. Air is exchanged through openings in the building envelope, usually through openable windows. Thus, the air supply is untreated regarding thermal and pollutant aspects. Therefore, indoor comfort is highly dependent on outdoor conditions.
Outdoor conditions are also the main factor influencing the air change rate in naturally ventilated buildings. Differences in temperature between interior and exterior and wind conditions are the main determinants. Control can only be enforced through the opening area of a building, making a defined air change rate infeasible. Natural ventilation is characterized by the low effort and costs required to integrate the required passive systems and its robustness throughout the usage of the building.

Extract systems (b) Extract systems employ mechanical systems for the forced removal of air. The induced underpressure leads to a reflow of fresh air through designated openings

[Hausladen and Tichelmann, 2012]. As in naturally ventilated buildings, the air supply is untreated upon entry. In contrast to naturally ventilated buildings, the air removal rate leads to the controllability of the supply airflow and the air change rate [Hausladen and Tichelmann, 2012], which, however, is still dependent on outdoor conditions.
Extract systems require more effort in planning and space than natural ventilation to integrate the extract system and ducts. Furthermore, energy is required to power the extract systems in operation.

Supply and extract systems (c) Supply and extract systems employ mechanical systems to remove and supply air. In contrast to natural ventilation and extract systems, supply air is usually pretreated to meet requirements regarding thermal comfort and indoor air quality. Supply and extract systems are highly controllable since mechanical systems induce all air movement. Supply and extract systems require a high effort in planning and space for systems and ductwork. Furthermore, maintenance and energy consumption require high expenses during the usage phase of a building [Liedl et al., 2011].

2.1.2 Topology of mechanical ventilation systems

Mechanical ventilation systems are systems to actively move and/or treat air. Its purpose is to provide acceptable indoor air quality and thermal comfort in occupied rooms [Hasager et al., 2021]. This is achieved by adjusting the temperature and humidity of the air and removing or diluting pollutants from the air [Hasager et al., 2021]. Mechanical ventilation systems employ fans to move the air and thus drive the air exchange; additional components to heat, cool, filter, humidify, and dehumidify are switched in series. In the following, the topology and the key components of mechanical ventilation systems will be examined.
The topology of mechanical ventilation systems can be differentiated into decentralized and centralized ventilation systems. A combination and fluent transition between centralized and decentralized systems are employed according to individual requirements.

Decentralised ventilation systems Decentralized ventilation systems are room-wise systems for extracting and supplying air [Hausladen and Tichelmann, 2012]. Decentralized systems are usually integrated into the building's facade and intake outdoor air via the facade requiring no ductwork[Liedl et al., 2011]. Decentralized ventilation systems allow for independent control and demand-driven supply in each room, with high user controllability [Hausladen and Tichelmann, 2012]. However, independence comes at the cost of high maintenance, acquisition, and installation costs. Especially the replacement of the filters leads to high costs during the usage phase and requires access to all rooms [Liedl et al., 2011].

Centralized ventilation systems Centralized ventilation systems for extracting and supplying air move and treat the air in stand-alone air handling rooms in the buildings, distribute the air through ducts, and supply and extract air via inlets and outlets [Hausladen and Tichelmann, 2012]. In centralized ventilation systems, ducts take a

considerable amount of space to achieve the required air change rate and require much energy for the transportation of air through the ductwork to compensate for pressure losses [Hausladen and Tichelmann, 2012]. Compared to decentralized ventilation systems, the central air treatment allows for equipment and filters with higher efficiency and simplifies maintenance for active components. However, the technical systems and ventilation ducts require a high hygienic standard, resulting in high maintenance costs for duct cleaning [Liedl et al., 2011]. The controllability of centralized ventilation systems is limited to setting the air change rate of the supply air to each individual zone using variable air volume units in the ductwork. Compared to decentralized systems, this leaves less space for user control.

2.1.3 Components of mechanical ventilation systems

In the following section, the treatment components making up a mechanical ventilation system are examined. These are heating and cooling, dehumidification and humidification, filtering, heat recovery, and recirculation.

Heating and Cooling The heating and cooling components pretreat the air to achieve comfortable supply air temperatures [Hausladen and Tichelmann, 2012]. Ventilation systems can either be the sole provider of heating and cooling to a room or in parallel to another heating/cooling system. The first requires the ventilation system to cover the total heating/cooling demand of the building. In most cases, this requires a significantly higher air change rate than hygienically required [Hausladen and Tichelmann, 2012] since air has a low heat storage capacity compared to its volume. If the ventilation system is operated in parallel to another heating/cooling system, the air change rate can be oriented on the hygienic demands and, thus, require considerably less volume flow. Air is heated in building ventilation systems through heating coils forming a heat register. Heating coils transfer heat to the air through forced convection [ASHRAE, 2020]. In order to heat the heating coils, either a medium such as steam, hot water, or refrigerant vapor is used, or they are heated directly with electricity. The heating register comprises 1 to 2 rows of steam coils or 2 to 4 rows of water coils with airflow at a right angle to the coils [ASHRAE, 2020]. Air is cooled in building ventilation systems through cooling coils forming a cooling register. Cooling coils absorb heat from the air through forced convection, and evaporation [ASHRAE, 2020]. Water glycol is used as the most common refrigerant in cooling coils [ASHRAE, 2020]. Depending on the required cooling capacity, multiple cooling coils are arranged in parallel and usually at a right angle to the airflow [ASHRAE, 2020]. In addition, water can be sprayed on the cooling coils to enhance the cooling capacity by evaporative cooling [ASHRAE, 2020].

Humidification and Dehumidification Humidification and dehumidification are the injection or removal of water vapor from air [ASHRAE, 2020]. Dehumidification is applied under ambient pressure using desiccant dehumidifiers or mechanical dehumidifiers. Desiccant dehumidifiers remove water vapor from the air by sorption methods [ASHRAE, 2020]. Either through non-regenerative processes with hygroscopic salts or in regenerative

processes applying substances with reversible water removal mechanisms such as silica gel [ASHRAE, 2020].
Mechanical dehumidifiers remove water vapor by passing the air over a cold surface below the dew point temperature of the air. Water vapor in the air condenses on the cold surface and can be drained by the system, and the air is reheated [ASHRAE, 2020]. In practice, this is usually done with a cooling and heating coil switched in series. Compared to regular cooling coils, in dehumidification, air is passed at lower velocities and with improved drainage of dew [ASHRAE, 2020].
Humidification is seldom applied in building ventilation systems due to the hygienic risk of fungal growth. If the required point of use, ultrasonic humidifiers are applied for comfort purposes (e.g., offices) and steam grid humidifiers if exact control is necessary (e.g., laboratories) [ASHRAE, 2020].

Filtering Filtering is meant to remove contaminants from the air, such as fine dust, viral and bacterial contaminants, fungal contaminants, and chemical pollutants [Hausladen and Tichelmann, 2012]. Filters are differentiated into particle and gaseous phase filters, referring to their filtering capacity. For particle filters, fiber filters are mostly applied. They remove particles by straining coarse particles with small openings and intercepting particles through van der Waals or electrostatic interactions [ASHRAE, 2020].
An alternative for particle pollutants are electronic filters, which ionize the particles in an electromagnetic field induced by wires conducting 6 to 25 kV electricity [ASHRAE, 2020]. The ionized particles are then captured by charged collection plates.
For gaseous phase filters, substances react with the target gases in the air. A common gaseous phase filter is activated carbon [ASHRAE, 2020]. In practice, multiple filters are combined to achieve the desired filtering effect. Since the filters significantly impact the system performance due to flow resistance, a regular exchange is required to upkeep performance and keep flow resistance low [ASHRAE, 2020].

Heat recovery Heat recovery systems aim to reuse the remaining heating or cooling energy in the extracted air [Hausladen and Tichelmann, 2012]. Heat recovery systems can be differentiated by their flow path, heat recovery, and moisture recovery efficiency.
Cross-flow heat exchanger systems employ crossing air ducts of extract and supply air with a high surface and contact area [Fuchs et al., 2008], resulting in a heat flow from the warm extract air to the cold supply air without mixing. Cross-flow heat exchangers recover up to 60% Counter-flow heat exchangers implement the same working principle as cross-flow heat exchangers; however, due to an optimized flow path and larger contact areas, heat recovery efficiency is up to 90% [Fuchs et al., 2008].
Run-around coils are used if the extract air outlet and supply air intake do not coincide. A brine circuit is a heat transfer medium between the extract and supply air. However, the system can only recover up to 50% of the heat due to inherent inefficiencies [Fuchs et al., 2008].
Rotary heat exchangers employ a rotating thermal mass, alternating between supply air and exhaust air stream. The system achieves a heat recovery efficiency of up to 80% and

a moisture recovery efficiency of up to 70% [Fuchs et al., 2008].
Most heat recovery systems usually have a bypass for mild weather conditions, in which heat recovery could increase energy demand [ASHRAE, 2020].

Recirculation Recirculation is mainly applied in air-based heating or cooling systems where the air change rate is higher than the hygienically necessary. In order to avoid losing heating or cooling energy through the exhaust, a fraction of the air is recirculated. The implementation can be differentiated into recirculation systems employing two separate circuits with a dedicated outdoor air circuit and systems with a single circuit with a variable fraction of recirculated air, so-called variable air volume systems [ASHRAE, 2017].
Dedicated outdoor air systems decouple the treatment and supply of outdoor air from the treatment and supply of recirculated air [ASHRAE, 2020]. The outdoor circuit's air volume is sized to meet hygienic demands.
Variable air volume systems control the temperature in a zone by varying the supplied air volume. The supply air temperature is usually determined building-wide by the most demanding zone [ASHRAE, 2020]. The supply volume of the outdoor air is sized in order to meet hygienic demands.

Air supply The air supply is the injection of pretreated air into a building zone. Air supply strategies can mainly be differentiated in mixing and displacement ventilation [Hausladen and Tichelmann, 2012].
Air supply by mixing ventilation injects the air with high velocity [Hausladen and Tichelmann, 2012]. Diffuser outlets ensure an even distribution of supply air within the building zone. The outlets are usually arranged in the ceiling or at high wall levels [Hausladen and Tichelmann, 2012]. Due to the high air velocities, draught might lead to discomfort if directly exposed to the air stream [Hausladen and Tichelmann, 2012]. Furthermore, distribution patterns might lead to lower air quality.
Air supply by displacement ventilation injects air at low velocity [Hausladen and Tichelmann, 2012]. Large outlets are arranged near the floors, creating a layer of fresh air at the bottom of the zone. After the air is heated by occupants or equipment, it rises, and stale air can be extracted at ceiling level [Hausladen and Tichelmann, 2012]. Low velocities and an even distribution of fresh air lead to high comfort and high air quality [Hausladen and Tichelmann, 2012].

2.1.4 Control and automation of mechanical ventilation systems

Control and automation in building ventilation systems are essential to optimize energy consumption, air quality, and operation costs. Measuring temperature, humidity, and pollutant concentration enables a demand-driven air supply [Coleman and Meggers, 2018]. Control and automation systems are working on three functional levels: field level, automation level, and management level [Hausladen and Tichelmann, 2012].
At the field level, sensors record information from devices about environmental conditions or occupation, and actuators perform tasks such as closing the sun blinds, opening a

valve, or switching on the light [Hausladen and Tichelmann, 2012].
At the automation level, interfaces and controls for sensors and actuators are arranged, implementing setpoints and rules from the management level. The management level enables user input and acts as a central monitoring system of all processes, implementing data visualization and alarming and storing historical data in databases [Hausladen and Tichelmann, 2012].
The communication between the different levels occurs via a BUS system, a digital protocol that assigns an individual address to each device by which it can be addressed. BUS systems transfer data either via a cable, which can also act as a power supply, or communicate via radio and enable battery operation [Hausladen and Tichelmann, 2012].
Demand-controlled ventilation refers to the continuous and automatic control of the air exchange rate to match the demand depending on occupancy, indoor pollution and thermal comfort [Mansson et al., 1997]. This represents a supplementary layer of control to conventional scheduled ventilation systems, informed by data from sensors [Coleman and Meggers, 2018]. Commonly, demand controlled ventilation systems employ sensors for humidity, carbon dioxide, or motion to approximate occupancy and assess the indoor air quality [Coleman and Meggers, 2018].

2.2 Indoor air quality within the concept of Indoor environmental quality

Indoor air quality can be classified within the concept of indoor environmental quality. Indoor environmental quality is assessing the conditions inside the building [Kubba, 2016]. It considers visual comfort, acoustic comfort, olfactory comfort, thermal comfort, and indoor air quality [Sarbu and Sebarchievici, 2013]. Its focus on human health, physiology, and psychology is common to all these aspects.
In the following section the individual aspects of indoor environmental quality are explored and delimited to indoor air quality. Figure 2.1 gives a visual summary of the primary sources of indoor air pollution, the types of contaminants they produce, and the resultant health impacts on building occupants, as well as, delimitating it from other aspects of indoor environmental quality.

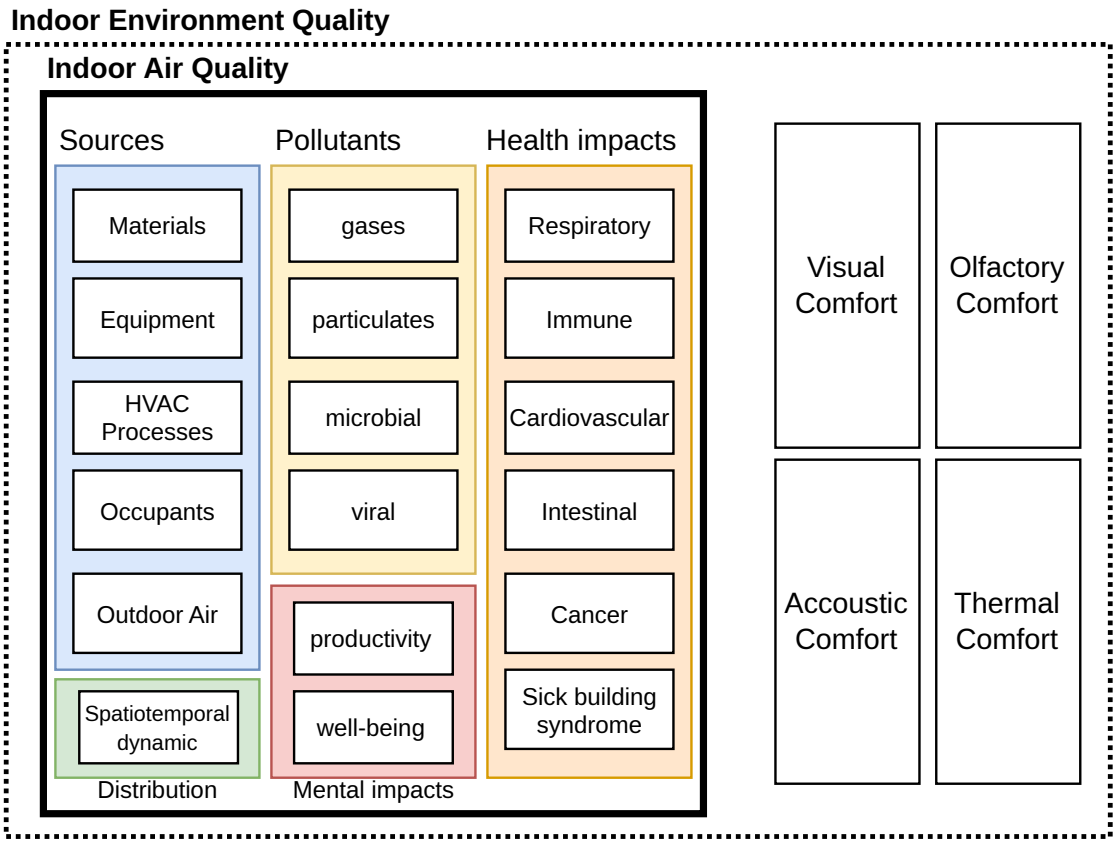


Figure 2.1: Visual classification of indoor air quality within indoor environmental quality

2.2.1 Visual comfort

Visual comfort is the effect and complex interrelation of light intensity, color, reflection, distribution, location, and daytime on human psychology and physiology [Hausladen and Tichelmann, 2012].
Light enables humans to orient themself and perceive their environment. It directly affects physiology by influencing the hormone release of the body depending on light intensity and color and directly affects alertness, body temperature, and fatigue. Humans can perceive light in intensities from 0.1 to 100.000 lux within the optical spectrum between 380nm to 780nm wavelength [Hausladen and Tichelmann, 2012]. The presence (and absence) of natural light in indoor environments is shown to have effects on occupant health, psychological well-being, productivity and sleep quality [Arif et al., 2016].
Indoor building materials heavily influence light perception through their physical reflectivity, transparency, translucency, and adsorption attributes. Furthermore, artificial or natural light sources impact visual comfort by their distribution, incidence angle, color spectrum, and light temperature [Hausladen and Tichelmann, 2012]. Especially glare impacts visual comfort in indoor environments, leading to fatigue and eye strain if light densities are unevenly distributed [Hausladen and Tichelmann, 2012].

2.2.2 Olfactory comfort

Olfactory comfort assesses the smell and whether it is pleasant or unpleasant to occupants. Since smell perception is individual and based on cultural and personal background, olfactory comfort is subjective to each person [Hausladen and Tichelmann, 2012].
Physiologically the sense of smell acts as a detection system for unhygienic or dangerous substances. The human sense of smell is capable of detecting gases at very low concentrations in the range of a few parts per billion and distinguishing over 10 000 different odors [Hausladen and Tichelmann, 2012]. "The perceived olfactory sensation depends not only on the pollutant source but also to a great extent, on the dilution degree with outside air." [Sarbu and Sebarchievici, 2013]
Air is a mixture of gases, with its major constituents being oxygen (21%) and nitrogen (79%) and many minor components depending on local sources. Indoor human effluents, human-emitted gases, and substances are one of the main odor sources [Hausladen and Tichelmann, 2012]. Human odor emissions are rated in olf for standardized activity and emission scenarios. The perceived odor is measured in decipol, standardized as one olf, diluted by 10 liters of fresh air per second [Hausladen and Tichelmann, 2012]. A maximum of 2.5 decipol is acceptable according to DIN 1946-2.

2.2.3 Accoustic comfort

Acoustic comfort is the absence of noise and providing an appropriate accoustic environment [Arif et al., 2016]. The perception of noise is subjective to each individual. Physiologically humans can hear frequencies from 16 to 20 000 Hz and a loudness of up to 120 - 140 dB without feeling pain [Hausladen and Tichelmann, 2012]. The perceived loudness is frequency dependent and is higher at medium pitches and lower at high and

low pitches [Hausladen and Tichelmann, 2012].
Sound is induced by vibrations of a body transmitted through a medium. Indoors the structure of the surrounding surfaces greatly influences the acoustic attributes through the reflection of sound by smooth, hard surfaces and absorption by porous, soft surfaces [Hausladen and Tichelmann, 2012]. The threshold for regularly occupied rooms lies between 25 - 55 dB [Hausladen and Tichelmann, 2012].

2.2.4 Thermal comfort

Thermal comfort is the human heat balance in relation to its surroundings, with the human body neither having excess heat nor heat loss [Hausladen and Tichelmann, 2012]. Thermal comfort is determined by indoor air temperature, mean radiant temperature of bordering surfaces, relative humidity of air, partial water vapours pressure, air velocity, thermal resistance of clothing as well as the degree of activity The PMV and PPD factors are a quantifying factor for thermal comfort [Sarbu and Sebarchievici, 2013]. The physiological factors are age, sex, and constitution of the individuals determining their heat balance point. Individual factors are dependent on experience and adaptability, ethnic influences, and psychosocial influences, such as the ability to control the environment [Hausladen and Tichelmann, 2012].
Physiologically humans are homeothermic at 37°C. Excess heat is compensated by sweating and heat loss by a decreased temperature in extremities, increased muscle activity, and increased metabolism [Hausladen and Tichelmann, 2012]. Physically the body stands in constant heat exchange with its surroundings. Heat is transferred through conduction by direct contact, longwave radiation between the skin and other visible objects, convection through airflow over the skin, and evaporation of water on the skin [Hausladen and Tichelmann, 2012].
The setting determines the clothing factor, the degree of activity, the food intake, and the heat balance point. Thermal comfort is time-dependent towards diurnal and annual rhythms due to the human hormone release and metabolism dependent on sunlight.
Generally, a room's surrounding surfaces and air temperature should not diverge more than 3 Kelvin [Hausladen and Tichelmann, 2012]. For office and residential use, operative temperature between 20 to 26 $°C$ is the range of temperatures where most people are satisfied. Furthermore, air velocities should stay below 0.19 m/s in summer and 0.16 m/s in winter with relative humidities between 30 - 65% [Hausladen and Tichelmann, 2012]. However, EN 15251 and ISO 7730 allows for a wider range of operative temperatures in free running buildings at elevated air speeds since research showed that elevated air velocity can increase the thermal comfort range by up to 3 Kelvin [Nicol and Humphreys, 2010].

2.2.5 Indoor air quality

Indoor air quality looks at the composition of the air regarding pollutants that negatively impact the health of occupants [Tham, 2016]. Due to its relevance for health, the maximum allowable concentrations for pollutants are often defined in regulations and

require monitoring.
In contrast to olfactory comfort, which focuses on the comfort impacts of odors, indoor air quality is focused on the health impacts of air pollutants. An unharmful substance can have a comfort impact by emitting a smell, and an odorless substance can have a health impact. Compared to olfactory comfort, which is difficult to quantify since odor intensity and perception are subjective sensations, indoor air pollutants are measurable, and their effects can be quantified.
The major groups of pollutants are organic gases, inorganic gases, particulates, microbial pollutants, and viral and bacterial infections [Tham, 2016]. Studies have shown that low indoor air quality leads to many different symptoms and health impacts. The severity and consequences of indoor air pollution are influenced by the type of pollutants present, their levels of concentration, the duration of exposure, and personal characteristics like age, general health, and physical condition [Berglund et al., 1992]. Common symptoms experienced by individuals in such environments include fatigue, eye irritation, headaches, and difficulty concentrating, collectively referred to as sick building syndrome [Berglund et al., 1992]. Extended exposure can also contribute to respiratory conditions and immune system responses, including asthma, particularly in sensitive populations like children and the elderly [Berglund et al., 1992]. Pollutants such as formaldehyde lead to inflammation of the skin and mucous membranes [Maroni et al., 1995], and certain groups of volatile organic compounds are carcinogenic [Berglund et al., 1992]. Further still sparsely researched, but likely effects are impacts to the cardiovascular system, the liver, kidney, and gastrointestinal tract [Berglund et al., 1992].
The pollutant sources and the building ventilation determine indoor air quality. Building materials, furnishing, and surface finishes passively emit pollutants to varying degrees [Tham, 2016]. Appliances are often active producers of pollutants, as heating and cooling systems enhance the distribution of pollutants [Tham, 2016] as do active air supply and extraction. Furthermore, occupants contribute to indoor air pollution through the resuspension of particles in the air through activity and through the emission of human bio effluents [Tham, 2016]. Lastly, pollutants are brought into the building from the outside through ventilation, infiltration through the facade, or other pathways [Tham, 2016].
Measures to improve indoor air quality are source control, by identifying and minimizing the sources; pathway control, by mitigation against pollutant pathways, reactive control, by initiating mitigation measures if thresholds are exceeded or predictive control, by initiating mitigative measures based on predictive models [Tham, 2016].

2.3 Indoor air pollutants

The following section provides an overview of the most prevalent and health critical pollutants affecting indoor air quality, detailing their chemical nature, health impacts, and sources. The selection is based on their presence in common standards as well as their occurence frequency in non residential indoor environments.

2.3.1 Carbon dioxide (CO_2)

Table 2.1: CO_2 exposure thresholds (absolute concentration)

duration	**1 h**	**8 h**	**24 h**	**long-term**
Canada	3504.0 ppm	-	-	-
China	-	-	1000.0 ppm	-
Europe	1000.0 ppm	-	-	-
Japan	-	-	-	1000.0 ppm
Singapore	-	1000.0 ppm	-	-
US	1000.0 ppm	-	-	800.0 ppm

Carbon dioxide is an inorganic gas and part of the atmosphere at a concentration of about 400 ppm, with an increasing trend due to the anthropogenic combustion of fossil fuels [Hasager et al., 2021]. Carbon dioxide is a non combustible and inert gas. Therefore, it is stable indoors and outdoors [Maroni et al., 1995]. It is colorless and odorless and cannot be detected by the human sense of smell [Maroni et al., 1995].
Chemically carbon dioxide acts as a simple asphyxiant on humans by displacing the oxygen from the air [Maroni et al., 1995]. High concentrations of carbon dioxide lead to a lack of oxygen and suffocation.
Its indoor sources are exclusively anthropogenic. The biggest factor is human breathing; exhalation concentrations are 100 times higher than inhalation concentrations at about 40 000 ppm compared to 400 ppm [Silbernagl and Despopoulos, 2003], leading to daily emissions of 1 kg carbon dioxide per person and day [Hasager et al., 2021]. Other indoor sources of carbon dioxide are combustion processes; however, in developed countries, this only plays an insignificant part [Maroni et al., 1995]. Carbon dioxide is often used as an indicator for other human bio effluents, due to its clear dependence on human activity [Hasager et al., 2021].
Indoors the typical concentration in occupied spaces is significantly higher than outdoors and is in a typical range between 400 to 2500 ppm [Ma et al., 2021] and sometimes up to 5000 ppm. Studies found indoor concentrations exceeding 5000 ppm if unvented interior combustion was used [Maroni et al., 1995].
Direct impacts of CO_2 exposure on the human body can be identified at concentrations of 10 000 to 15 000 ppm, leading to physiological effects such as changes in blood pH and the carbon dioxide content in the blood [Maroni et al., 1995]. At concentrations between 15 000 and 30 000 ppm, breathing becomes faster and more difficult, 30 000 to 60 000

lead to dizziness and nausea [Maroni et al., 1995]. Exposure to concentrations above 60 000 ppm leads to stupor and death [Maroni et al., 1995].
Besides short-term exposure to high-level carbon dioxide concentrations, the effects of long-term exposure to lower concentrations have been extensively studied. According to Hasager et al. [2021], concentrations between 1000 and 10,000 ppm lead to occupants experiencing a loss of concentration, fatigue, and headaches. Ma et al. [2021] reported that continuous exposure to concentrations above 1000 ppm leads to physiological symptoms such as headaches, fatigue, and respiratory tract symptoms, as well as neurophysiologic symptoms such as impaired concentration, decision making, cognitive performance deficits, and fatigue. The study by Satish et al. [2012] found significant decrements in decision-making performance at concentrations of 2500 ppm in a study involving 22 participants exposed to varying CO_2 concentrations. Conversely, a study researching the effects of elevated CO_2 concentrations on decision making and cognitive functions in astronaut-like subjects found no correlation between test-subject performance and concentrations up to 5000 ppm [Scully et al., 2019].
Limits enforced for carbon dioxide date back until 1858 by Pettenkofer, defining a threshold of 1000 ppm for occupied rooms and 700 ppm for sleeping rooms [Abdul-Wahab et al., 2015]. This threshold is valid until today and is advised in the US, Europe, China, Japan, and Singapore [Abdul-Wahab et al., 2015]. The US environmental protection agency advises a slightly lower limit of 800 ppm for continuous exposure [Abdul-Wahab et al., 2015].

2.3.2 Carbon monoxide (CO)

Table 2.2: CO exposure thresholds

duration	**15 min**	**30 min**	**1 h**	**8 h**	**long-term**
Australia	-	-	-	9.0 ppm	-
Canada	-	-	29.0 ppm	11.0 ppm	-
China	-	-	10.0 ppm	-	-
Europe	100.0 ppm	60.0 ppm	30.0 ppm	10.0 ppm	-
Japan	-	-	20.0 ppm	-	10.0 ppm
Singapore	-	-	-	10.0 ppm	-
US	-	-	35.0 ppm	9.0 ppm	-

Carbon monoxide is an inorganic gas generated in incomplete combustion [Hasager et al., 2021]. It is non-combustible and inert and is therefore stable in indoor and outdoor conditions [Maroni et al., 1995]. Carbon monoxide is a colorless, odorless, and tasteless gas that humans cannot recognize. Chemically carbon monoxide reacts with the hemoglobin cells in the blood to carboxyhemoglobin [Maroni et al., 1995]. It is more than 200 times more effective in reacting with the hemoglobin as oxygen, reducing the oxygen concentration in blood [Maroni et al., 1995].
Indoor sources of carbon monoxide are incomplete combustion processes such as unvented

furnaces, smoking, and exhaust fumes [Maroni et al., 1995]. Outdoor sources include exhaust fumes from combustion engines and natural sources such as wildfires [Maroni et al., 1995].
The ambient concentration is about 1 ppm [Hasager et al., 2021]. Urban spaces have significantly higher concentrations of 10 - 20 ppm [Hasager et al., 2021]. If no indoor combustion sources are present, indoor concentration levels correlate closely with outdoor concentration levels [Hasager et al., 2021]. Indoors studies discovered a mean concentration in office buildings of 3.6 ppm [Maroni et al., 1995].
Exposure to carbon monoxide impacts health at concentrations above 35 ppm. Occupants exposed to a concentration level of 35 ppm for a working day experience dizziness and headaches due to the reduced oxygen absorption capacity [Maroni et al., 1995]. Exposure for 2 to 3 hours of 100 ppm leads to headaches, at 200 ppm to loss of judgment. Exposing occupants to 800 ppm for 45 minutes leads to dizziness, nausea, and convulsions [Maroni et al., 1995]. Carbon monoxide concentration levels above 1000 ppm lead to death quickly. Exposure to 1600 ppm leads to death within 2 hours, 3200 ppm within 30 minutes, and 12 800 ppm within 3 minutes [Maroni et al., 1995]. Even below 35 ppm, long-term exposure to elevated carbon monoxide concentration levels has detrimental effects on the heart, lungs, and nervous system [Ma et al., 2021]. Exposure limits to carbon monoxide are closely regulated in most countries and limit continuous exposure to a maximum of 8 to 10 ppm [Abdul-Wahab et al., 2015]. For shorter periods, Europe set a maximum of 15 minutes of 87 ppm, 52 ppm for 30 minutes, and an hourly exposure maximum of 25 to 30 ppm in Europe, the US, and Canada [Abdul-Wahab et al., 2015].

2.3.3 Particulate matter (PM)

Table 2.3: PM_{10} exposure thresholds

duration	**24 h**	**long-term**
Australia	-	0.09 mg/m³
China	0.15 mg/m³	-
Europe	0.5 mg/m³	0.2 mg/m³
Singapore	-	0.15 mg/m³
US	0.15 mg/m³	0.05 mg/m³

Particulate matter (PM) is a mixture of organic and inorganic compounds [Maroni et al., 1995]. It can be differentiated in coarse particulate matter (PM_{10}) with an aerodynamic diameter above 2.5 μm and fine particulate matter ($PM_{2.5}$) with an aerodynamic diameter below 2.5 μm [Maroni et al., 1995]. The properties of the pollutants in these groups are very heterogeneous.
The pathways and effects on humans depend majorly on their size. Fine $PM_{2.5}$, below 2.5 μm, is inhalable, penetrates the lungs, and may even enter the bloodstream [Maroni et al., 1995]. Coarse PM $_{10}$ above 2.5 μm is respirable but is mostly stopped from penetrating the lungs in the nose [Maroni et al., 1995]. PM_{10} has a variety of sources, both of natural

and anthropogenic origin. Natural sources include pollen, dust mites, mold, and viral and bacterial matter [Maroni et al., 1995]. Anthropogenic sources include combustion, cooking, smoking, and particle resuspension through human activity [Maroni et al., 1995]. In office buildings, the indoor-outdoor ratio is usually close to or below 1; in residential buildings, the indoor concentration regularly exceeds outdoor concentration [Maroni et al., 1995].

The impact of $PM_{2.5}$ on human health depends on the pollutant source. Soot from combustion processes is mutagenic [Maroni et al., 1995]. Polycyclic aromatic hydrocarbons, generated through combustion processes, are carcinogenic [Maroni et al., 1995]. Furthermore, $PM_{2.5}$ may cause inflammation of mucous membranes and immune reactions.
For fine $PM_{2.5}$, a continuous exposure below 10 $\mu g/m^3$ is enforced in Europe and 15 $\mu g/m^3$ in the US [Abdul-Wahab et al., 2015]. The maximum exposure for 24 hours is 25 $\mu g/m^3$ in Europe and 60 $\mu g/m^3$ in the US [Abdul-Wahab et al., 2015]. Continuous exposure for coarse $PM_{2.5}$ is limited to 200 $\mu g/m^3$ in Europe, and 50 $\mu g/m^3$ in the US [Abdul-Wahab et al., 2015]. The maximum exposure for 24 hours is set to 150 $\mu g/m^3$ in the US and 500 $\mu g/m^3$ in Europe [Abdul-Wahab et al., 2015].

Table 2.4: $PM_{2.5}$ exposure thresholds

duration	**1 h**	**8 h**	**24 h**	**long-term**
Canada	0.1 mg/m³	0.04 mg/m³	-	-
Europe	-	-	0.02 mg/m³	0.01 mg/m³
US	-	-	0.06 mg/m³	0.02 mg/m³

2.3.4 Nitrogen dioxide (NO_2)

Table 2.5: NO_2 exposure thresholds

duration	**1 h**	**24 h**	**long-term**
Canada	0.48 ppm	-	0.1 ppm
China	0.24 ppm	-	-
Europe	0.4 ppm	0.15 ppm	-
US	-	-	0.1 ppm

Nitrogen dioxide is an inorganic gas and is the most common from the group of NOx gases [Maroni et al., 1995]. Its chemical attributes are its water solubility, and it acts as an oxidizing agent [Maroni et al., 1995]. Nitrogen dioxide is red to brown and has an acrid odor [Maroni et al., 1995]. In humans, the oxidizing effect causes severe irritation and inflammation of the mucous membranes [Maroni et al., 1995]. Nitrogen dioxide is mainly generated in anthropogenic processes such as in high-temperature combustion processes by reactions between oxygen and nitrogen in the air [Maroni et al., 1995].

Combustion engines in vehicles are the primary source of nitrogen dioxide in urban areas. If no indoor combustion processes are performed, the main source of indoor nitrogen dioxide comes from the outside, with a typical indoor-outdoor ratio between 0.5 to 1 [Maroni et al., 1995]. If indoor combustion processes are present indoor concentrations significantly exceed outdoor concentrations by a factor of 2 to 5 [Maroni et al., 1995].
Nitrogen dioxide has severe health impacts at minimal concentrations. Impaired breathing and changes in pulmonary function are detectable at exposure to concentrations of 2 ppm for healthy adults and 0.5 ppm for vulnerable persons as asthmatics [Ma et al., 2021]. Children are especially susceptible to nitrogen dioxide exposure and show responses at levels as low as 50 to 300 ppb [Ma et al., 2021]. Continuous exposure to elevated nitrogen dioxide concentrations increases the susceptibility to infections, cause lasting lung irritation and damage, and leads to biochemical alterations [Ma et al., 2021].
Regulations define a threshold of maximum continuous exposure of 0.1 ppm to 0.15 ppm. Short-term exposure for an hour is set to 0.8 ppm in Germany, 0.48 ppm in Canada, and 0.4 ppm in Europe [Abdul-Wahab et al., 2015].

2.3.5 Ozone (O_3)

Table 2.6: O_3 exposure thresholds

duration	**1 h**	**8 h**	**long-term**
Australia	0.11 ppm	-	-
Canada	0.12 ppm	-	-
China	0.08 ppm	-	-
Europe	0.09 ppm	0.06 ppm	-
Singapore	-	0.05 ppm	-
US	0.12 ppm	0.08 ppm	0.05 ppm

Ozone is a highly reactive inorganic gas with a half-life of 3 days. It is colorless and odorless and acts as an oxidizing agent [Maroni et al., 1995]. Ozone chemically reacts with tissues and membranes of humans, causing irritation and inflammation [Maroni et al., 1995]. The main source of ozone is its generation in the stratosphere at about 25km height at wavelengths below 240 nm [Maroni et al., 1995]. Anthropogenic sources are the use of UV-lighting, for example, in air cleaners, photocopying machines, or laser printers [Maroni et al., 1995].
Daily outdoor concentrations are at a mean of 60 ppb with a peak of 180 ppb in the afternoon [Maroni et al., 1995]. Indoor concentrations are below outdoor concentrations. If no indoor sources are present, the indoor-outdoor ratio is between 0.1 to 0.25; at higher air change rates, the indoor-outdoor ratio can increase to 0.8 [Maroni et al., 1995].
Ozone has detrimental health impacts at levels as low as 80 ppb. Exposure to 80 ppb for one hour reduces pulmonary function in children and young adults [Ma et al., 2021]. Concentrations of 100 ppb exposed for one hour lead to eye, nose, and throat irritation, chest discomfort, coughing, and headache [Ma et al., 2021]. Humans exposed to ozone

levels above 100 ppb suffer from decreased performance and asthmatic attacks [Ma et al., 2021]. Standards and laws worldwide define a maximum continuous exposure between 50 and 60 ppb [Abdul-Wahab et al., 2015]. The hourly maximum is set to 120 ppb in the US, 90 ppb in Europe and 80 ppb in China [Abdul-Wahab et al., 2015].

2.3.6 Formaldehyde (HCHO)

Table 2.7: HCHO exposure thresholds

duration	**30 min**	**1 h**	**8 h**	**long-term**
China	-	0.08 ppm	-	-
Denmark	-	-	-	0.12 ppm
Europe	0.08 ppm	-	-	-
Japan	0.08 ppm	-	-	-
Singapore	-	-	0.1 ppm	-

Formaldehyde is an organic pollutant. It is the most present and health relevant pollutant from the group of the aldehydes [Maroni et al., 1995]. It is colorless and has a pungent odor; the human sense of smell can detect it at concentrations above 80 ppb [Maroni et al., 1995]. Formaldehyde is a reactive gas; in sunlight and with other urban pollutants, formaldehyde reacts to carbon dioxide with a half-life of 50 minutes [Maroni et al., 1995]. Formaldehyde is mainly emitted from anthropogenic sources and processes. Indoors the most common sources are adhesives and glues in furniture and building materials and lacquers and varnishes [Maroni et al., 1995]. Materials and surfaces gradually emit formaldehyde, and resins in wood products emit over several years after the production [Maroni et al., 1995]. Formaldehyde is also used as disinfectant, resulting in elevated concentrations in medical buildings [Maroni et al., 1995].
The natural outdoor concentration of formaldehyde is 0.8 ppb, and urban areas show an elevated level with a daily mean concentration of 4 to 8 ppb and 40 to 80 ppb during peak traffic times and smog events [Maroni et al., 1995]. Indoor concentrations of formaldehyde exceed outdoor concentrations. Furthermore, no significant correlation between indoor and outdoor correlation has been found [Maroni et al., 1995]. Studies of the German building stock showed that 94% of the buildings have formaldehyde concentrations below 100 ppb [Maroni et al., 1995].
Exposure to formaldehyde has significant health effects above levels of 400 ppb. Starting at concentrations of 400 ppb, prolonged exposure irritates the eyes and throat [Maroni et al., 1995]. Exposure to 5 ppm causes immediate watering of the eyes. Concentrations above 30 ppm lead to life-threatening inflammation, edema, and pneumonia [Maroni et al., 1995]. Formaldehyde concentrations above 100 ppm lead to death [Maroni et al., 1995].
Studies have shown that long-term exposure to elevated formaldehyde concentrations leads to the development of asthma and has potentially mutagenic and carcinogenic effects [Maroni et al., 1995]. Regulations to restrict the exposure to formaldehyde set a

threshold of 80 ppb for a maximum of 30 minutes in Europe or 60 minutes in Japan and China [Abdul-Wahab et al., 2015].

2.3.7 Sulfur dioxide (SO_2)

Table 2.8: SO_2 exposure thresholds

duration	**1 h**	**24 h**	**long-term**
Australia	0.57 mg/m³	-	0.6 mg/m³
China	0.5 mg/m³	-	-
US	0.35 mg/m³	0.36 mg/m³	0.08 mg/m³

Sulfur dioxide is a colorless, inorganic gas with a pungent odor, sensible at concentrations above 500 ppb. It is a highly reactive gas that oxidizes to harmful acid aerosols [Maroni et al., 1995]. Its main source is the combustion of fossil fuels, most relevant in combustion engines of vehicles.
Outdoor concentrations in urban areas are below 40 ppb on a daily mean. However, peaks can go as high as 750 ppb [Maroni et al., 1995]. Indoor concentrations are lower than outdoor concentrations if no internal combustion of fossil fuel occurs. Sulfur dioxide has a short half-life interior due to its high reactivity. The typical indoor-outdoor ratio is between 0.1 and 0.6 [Maroni et al., 1995]. Exposure to sulfur dioxide is especially critical to vulnerable groups such as asthmatics and children. Adolescent asthmatics react to concentrations of as low as 40 ppb, adult asthmatics at exposure from 130 ppb [Maroni et al., 1995]. Long-term exposure to elevated levels induces chronic bronchitis [Maroni et al., 1995]. Regulations limit the allowable continuous exposure limit to 12 ppb in Europe, 30 ppb in the US, and 20 ppb in Australia [Abdul-Wahab et al., 2015].

2.3.8 Volatile organic compounds (VOC)

Table 2.9: VOC exposure thresholds

duration	**1 h**	**8 h**	**long-term**
Australia	0.19 mg/m³	-	-
China	-	0.23 mg/m³	-
Singapore	-	-	0.5 mg/m³

Volatile organic compounds are a group of organic pollutants with a low boiling point between 50°C and 260°C. Over 900 volatile organic compounds (VOC) are relevant indoors [Maroni et al., 1995]. Almost any artificial material releases volatile organic compounds. Main indoor sources are cleaners, paints and lacquers, adhesives, cosmetics, furniture, and building materials [Maroni et al., 1995]. Volatile organic compounds have a narcotic effect and suppress the human central nervous system [Maroni et al., 1995].

If not individual volatile organic compounds are measured, concentration is given in a VOC equivalent, as summed mass concentration of all volatile organic compounds [Maroni et al., 1995]. Indoor concentrations significantly exceed outdoor concentrations due to the presence of artificial materials. A study in 83 office buildings found a median concentration of 84 ppb VOC with 90% below 330 ppb [Maroni et al., 1995]. The effects and exposure levels of individual volatile organic compounds are highly heterogeneous. Most common causes of elevated short-term exposure are irritation of the eyes, the respiratory tract and the skin. Furthermore, volatile organic compounds can cause psycho-physiological effects and result in narcotic effects, fatigue, headaches, and concentration problems [Maroni et al., 1995]. According to Tham [2016], volatile organic compounds are one of the main sources of the sick building syndrome.
Regulations limit the allowable VOC concentration to an average of 230 ppb over 8 hours in China; continuous exposure is limited to 500 ppb in Singapore [Abdul-Wahab et al., 2015].

2.4 Summary

To summarize the fundamentals section of this dissertation, the terms and use of building ventilation were recapped. Building ventilation is the controlled movement of air in indoor environments to supply fresh and unpolluted air and extract old and polluted air. In general, there are various approaches to building ventilation; the two major categories are natural and mechanical ventilation. However, in most nonresidential buildings with focus on open-office and classroom typologies, a hybrid approach, combining natural and mechanical ventilation, is the rule.
Furthermore, the section took a deeper look at mechanical ventilation systems, generally defined as systems that use fans to move the air and include optional components for heating, cooling, dehumidification, humidification, filtering, heat recovery, and recirculation. It was also noted that mechanical ventilation systems can generally be differentiated into systems that only extract air (extract systems) versus systems that extract air and supply fresh air (supply and extract systems). A further differentiation was identified regarding centralized or decentralized mechanical ventilation systems. The former uses a central device to supply the whole building with air and extract air. The treated air and exhaust air is moved in ducts to the individual zones. Decentralized systems, on the other hand, are located within the zones they supply air to and usually require no or minimal duct work. Especially in central ventilation systems, the ventilation control and automation system plays a major role in automating and optimizing energy consumption, air quality, and operation costs.
Further on in the theory section, the term indoor air quality and indoor air pollutants was defined and classed within the overarching field of indoor environmental quality (IEQ). Indoor environmental quality considers visual comfort, acoustic comfort, olfactory comfort, thermal comfort, and indoor air quality. Indoor air quality was delimited from olfactory comfort, which considers the subjective perception of smell. In contrast, indoor air quality is affected by physically measurable pollutants such as carbon dioxide, carbon monoxide, particulate matter, nitrogen dioxide, ozone, formaldehyde, sulfur dioxide, and volatile organic compounds to name the most relevant.
In the last part of the theory section, the most relevant indoor air pollutants were listed and analyzed regarding their chemical nature and its effect on humans, their intake pathways, their natural and anthropogenic sources, and the limits and thresholds enforced in different regulatory systems worldwide. An in-depth analysis of these thresholds was made for each pollutant, comparing the limits enforced in various global regions and examining these thresholds combined with the defined exposure times.
It is noted that the effect of pollutants can range from irritation and inflammation of the respiratory system to long-term effects such as asthma and even carcinogenic effects, and it was identified that it is therefore crucial to closely monitor and control individuals' indoor air pollutant exposure to ensure health and well-being of occupants in indoor spaces.

3 State of the Art

The following chapter reviews current research, which examines the state of indoor air pollution in non residential buildings, measuring indoor air pollutants, virtual sensing of indoor air pollutants using BMS data and demand-controlled ventilation. The chapter is finalized by describing the contribution this dissertation will make in relation to the reviewed literature.

3.1 Indoor air pollutants in nonresidential buildings

The first section of this chapter examines the state of indoor air pollutant concentration in nonresidential buildings by looking at recent studies that researched and monitored indoor air pollutants in these environments. The section looks at the pollutants measured by individual studies, the measurement setup, the determinants of indoor air pollutants, the variations within and between buildings, and typologies.

3.1.1 Measured pollutants

Fifteen studies monitoring indoor air pollutants in nonresidential buildings were assessed (see table 3.1). Two-thirds of these studies have been published after 2017, showing the emergence of indoor air pollutant monitoring, with studies in Asia, Africa, Australia, Europe, and North America. In the referenced research, $PM_{2.5}$ emerged as the most frequently analyzed pollutant, with 14 of the 15 studies focusing on it. Volatile organic compounds were the subject of investigation in 8 out of 15 studies, while CO_2 was measured in 7 out of 15. Carbon monoxide and nitrogen dioxide were assessed in 5 of the 15 studies each. Additionally, ozone and sulfur dioxide were evaluated in two studies. Other pollutants, such as nitrogen oxide and fungi, were only included in a single publication's scope.

Due to the large variety of indoor air pollutants, studies mainly measure a subset of all pollutants. The selection in given studies is influenced by the availability of measuring equipment and the importance of a pollutant in a given environment, as found in previous research.

The key findings regarding pollutant concentrations of the individual studies are as follows. The study by [Ma et al., 2021] revealed that the accumulation of carbon monoxide and radon is primarily confined to specific areas, notably kitchens equipped with gas stoves and basements. In contrast, these pollutants are generally negligible in non-residential buildings. Zhang et al. [2021] underlines the negligible influence of carbon monoxide in nonresidential buildings; the results indicate concentrations below the measuring threshold. Additionally Irga and Torpy [2016] found carbon monoxide, nitrogen oxide,

Table 3.1: Examined pollutants in research

Study	$PM_{2.5}$	PM_{10}	VOC	CO_2	NO_2	CO	O_3	SO_2	NOx	NH_3
[Zhang et al., 2021]	x	x	x	x	x	x	x	x		
[Kim et al., 2019]	x	x		x						
[Saraga et al., 2017]	x	x								
[Irga and Torpy, 2016]	x	x	x	x	x	x		x	x	
[Montgomery et al., 2015]	x	x	x	x						
[Ha et al., 2020]			x	x		x				x
[Kang and Hwang, 2016]		x	x			x				
[Mendoza et al., 2021]	x									
[Tiele et al., 2018]	x	x		x	x	x				
[Spinazzè et al., 2020]	x		x		x		x			
[Li et al., 2018]	x									
[Saini et al., 2020]	x	x								
[Challoner and Gill, 2014]	x				x					
[Challoner et al., 2015]	x				x					
[Ahn et al., 2017]	x	x	x	x						

volatile organic compounds, and sulfur dioxide levels insignificant in 11 office buildings with mechanical ventilation systems and natural ventilation. Furthermore, Irga and Torpy [2016] measured indoor fungi concentrations, with results indicating a seasonal variability but being continuously below health-relevant thresholds. Irga and Torpy [2016] and Challoner and Gill [2014] independently found an indoor-outdoor ratio for nitrogen oxides close to one, indicating that the ground level of buildings near traffic-intensive streets might be problematic. Apart from this exception, Irga and Torpy [2016] found that nitrogen dioxide concentrations are far below their respective threshold and are barely influenced by occupant activity or air change rate. The most problematic pollutants in nonresidential buildings are $PM_{2.5}$, CO_2 and VOC. For $PM_{2.5}$, Challoner and Gill [2014] identified overruns in 10% of the measured time and Irga and Torpy [2016] ascertaining a seasonal dependence with peak exceedances in October. Moreover, the study by Montgomery et al. [2015] found that carbon dioxide levels frequently surpassed the 1000 ppm threshold in buildings with natural ventilation. This was attributed to an inadequate rate of air exchange. Mechanical ventilation systems effectively control volatile organic compounds and formaldehyde but can reach problematic concentrations if the ventilation system is switched off [Montgomery et al., 2015].

3.1.2 Measurement setup

In total, 67 nonresidential buildings and 194 rooms were examined for indoor pollutant concentrations in the reviewed studies. Sixty-two buildings are commercial and eight educational. 9 of 15 studies reviewed a single room in a single building. More than one building and more than one room were examined respectively in four studies. More than one room per building for more than one building was only assessed in [Spinazzè et al., 2020; Mandin et al., 2017], examining 148 rooms in 37 buildings. 11 out of 15 studies provide information regarding the type of room where indoor air quality is measured. For these, 70% of the studies examine indoor air quality in open office spaces and 10% either cafeteria, workshop, or laboratory spaces. 8 out of 15 studies provide detailed information on the HVAC system used in the respective buildings. Each of these studies examines at least one building with a full mechanical ventilation system; this is compared to naturally ventilated buildings in six studies.

More than half of the examined studies perform measurements for a week or less. With two studies measuring below one day [Tiele et al., 2018; Li et al., 2018]. One-third of all studies measure indoor air quality for a month or more, with Mendoza et al. [2021] and Irga and Torpy [2016] measuring for a full year.

The most assessed month is August, covered by eight studies, followed by six studies for July, June, and April. November and December are least represented and are only examined in the full-year studies by Mendoza et al. [2021] and Irga and Torpy [2016]. The

Table 3.2: Measurement Setup of examined studies

Study	Building	Rooms	MV	NV	Typology
[Zhang et al., 2021]	1	1			Office
[Kim et al., 2019]	1	2			Office
[Saraga et al., 2017]	1	1	x		
[Irga and Torpy, 2016]	11	11	x	x	Office
[Montgomery et al., 2015]	1	1	x	x	Office
[Ha et al., 2020]	1	16			Office
[Kang and Hwang, 2016]	1	1			
[Mendoza et al., 2021]	1	1	x		
[Tiele et al., 2018]	1	1	x		Lab.
[Spinazzè et al., 2020]	37	148			Office
[Li et al., 2018]	1	1	x		Workshop
[Saini et al., 2020]	1	1			Cafeteria
[Challoner and Gill, 2014]	6	6	x	x	Office
[Challoner et al., 2015]	2	2	x	x	Office
[Ahn et al., 2017]	1	1		x	

second least studied months are October, January, February, and March, assessed by three studies. In summary, the winter and autumn months are significantly underrepresented in indoor air quality research, and most research is done during summer.
The sampling interval of the sensors defines the temporal resolution. The median sampling interval for the ten studies specifying it is 5 minutes. Only Li et al. [2018] and Mendoza et al. [2021] are gathering samples below one minute. Saraga et al. [2017] only samples once during 24 hours. Most studies employ a sampling scheme between 1 and 10 minutes; research on applicable temporal resolutions for indoor air pollutant measurements is missing, studies suggest high fluctuations on a sub-minute level [Ciuzas et al., 2015].
Ciuzas et al. [2015] performs an in-depth study on the temporal variability and distribution of fine particles of different diameters and from different sources. The study identified a fast increase and decrease for especially critical sources such as cigarette smoking and candle burning with peaks reached after a few minutes. Furthermore, Ciuzas et al. [2015] found a significant variation of instantaneous concentrations due to airflow patterns.
Therefore, sampling intervals above a minute increase the risk of leaving critical peaks in pollutant concentration undetected, thus, underestimating the health impact. Furthermore, prediction and control systems require high temporal resolutions to perform optimally.
Li et al. [2018] identifies optimal prediction results of an artificial neural network at a temporal resolution of one second and a window size of 30 seconds. Increasing the window size results in overfitting of the prediction, and a lower window size leads to insufficient training data and, thus, to a decrease in prediction accuracy. The spatial distribution in a single room is only examined by Li et al. [2018], distributing eight measurement nodes in a woodworking workshop in order to derive the spatiotemporal distribution of fine dust. All other studies gather samples exclusively in one location in each room.

3.1.3 Determinants of indoor air pollutants

Most studies perform additional measurements alongside indoor air pollutants to identify determinants and influences on indoor air pollutants and evaluate other aspects of indoor environmental quality. Measuring temperature and humidity is most common and performed in 8 out of 15 studies. Illumination is assessed in two studies and other factors, such as noise, pressure, and wind, respectively, in a single publication.
Identifying the determinants of indoor air pollutants is crucial in preventing air pollution events through active and passive mitigation measures and source control. Spinazzè et al. [2020] identified the main factors impacting indoor pollutants to be the building characteristics, equipment, indoor climate, occupant and cleaning activity, the building location and the proximity of traffic, and the country. The following section will examine the factors determining indoor air pollutants, differentiated by indoor climate, occupancy, outdoor environment, equipment, and HVAC system.

Influence of indoor climate Several studies showed a significant impact of the indoor climate (temperature, relative humidity) on indoor pollutants and their distribution.

Table 3.3: Measured determinants of indoor air pollutants

Study	Temp.	Hum.	Light	Noise	Pres.
[Zhang et al., 2021]	x	x			
[Kim et al., 2019]	x	x			
[Saraga et al., 2017]					
[Irga and Torpy, 2016]	x	x			
[Montgomery et al., 2015]					
[Ha et al., 2020]	x	x			
[Kang and Hwang, 2016]	x	x			
[Mendoza et al., 2021]					
[Tiele et al., 2018]	x	x	x	x	
[Spinazzè et al., 2020]					
[Li et al., 2018]					
[Saini et al., 2020]	x	x			
[Challoner and Gill, 2014]					
[Challoner et al., 2015]				x	x
[Ahn et al., 2017]	x	x	x		

Zhang et al. [2021] identified a strong correlation between temperature, ozone, and $PM_{2.5}$. Zhang et al. [2021] attributed a Pearson correlation coefficient of 0.8 for ozone to the simultaneously increased solar irradiation and, thus, ozone generation. Fine $PM_{2.5}$ exhibited a correlation coefficient of 0.88, attributable to the simultaneous relative humidity decrease, which causes increased $PM_{2.5}$ distribution. Thus, relative humidity correlated negatively with fine $PM_{2.5}$ (R=-0.88) and ozone (R=-0.78). A strong correlation has been identified between humidity and carbon dioxide levels (R=0.78) and volatile organic compound concentration (R=0.72) [Zhang et al., 2021].
Spinazzè et al. [2020] ascertained a strongly positive correlation between volatile organic compounds and indoor temperature and relative humidity due to increased material emissions at higher temperatures and increased chemical reactions and volatile organic compound creation at higher relative humidity levels [Spinazzè et al., 2020].
An et al. [2010] showed that in buildings with floor radiant heating systems, the floor's temperature is majorly correlated to volatile organic compounds and formaldehyde emissions. Increased temperatures promote pollutant emissions from the floor structure. In Kang et al. [2010], this effect has been attributed to the adhesives in floor materials, which are highly temperature sensitive.

Influence of occupancy Jiang et al. [2020] examined the effect of occupancy on volatile organic compounds, ozone, carbon dioxide, and $PM_{2.5}$. Ozone concentrations were unaffected by space occupancy; $PM_{2.5}$ exhibited increased daytime concentrations, however, with no clear relation to occupancy [Jiang et al., 2020]. The effects of occupancy on

volatile organic compounds were more diverse; some pollutants, such as formaldehyde and benzene, showed no correlation to occupancy and showed constant emissions only correlated to materials present. Other volatile organic compounds were strongly correlated to occupancy since sources were human-associated, either as human bio effluents or used in care products [Jiang et al., 2020].
Spinazzè et al. [2020] examined the impact of occupants' activity patterns on volatile organic compounds and found that especially using body-hygiene products (hair sprays, deodorants) majorly impacted volatile organic compound concentration.
Kim et al. [2019] examined the indoor $PM_{2.5}$ concentration under consideration of different filter materials, occupation, and occupant activity. Activities for door or window opening, walking, or standing up were monitored and counted for a 10-minute time horizon. Carbon dioxide was found to be moderately correlated (Pearson coefficient = 0.44 - 0.49) to the number of occupants and uncorrelated to the occupants' activity. $PM_{2.5}$ is not correlated to the number of occupants but is slightly related to the activity (Pearson coefficient = 0.17 - 0.37) [Kim et al., 2019].

Influence of outdoor environment Several studies examined the indoor/outdoor relationship of different pollutants and the influence of the surroundings of the building. Mendoza et al. [2021] examined the correlation between in- and outdoor $PM_{2.5}$ concentration in a full-year study. Sensors are located on the rooftop, close to the ventilation system intake, in the air handling room, and office space. Correlations are examined for major pollution events, winter inversion, wildfire, and fireworks [Mendoza et al., 2021]. Depending on the type of pollution, the indoor and outdoor pollution correlated to a certain degree, depending on air change rate and filter efficiency. The highest correlations (R^2=0.99) were measured for wildfire events, as the particles could pass unfiltered through the air handling unit. Higher filtration efficiency could be achieved for fireworks (R^2=0.86, 0.66) and winter inversion (R^2=0.84) [Mendoza et al., 2021].
Saraga et al. [2017] identified a strong indoor/outdoor correlation of $PM_{2.5}$ during HVAC operation (Pearson correlation = 0.94) and a neglectable correlation during times without air change (Pearson correlation under 0.3). Thus, Saraga et al. [2017] ascertains a neglectable role of $PM_{2.5}$ ingress through infiltration and the ventilation system as the main impactor.
Irga and Torpy [2016] examined the indoor/outdoor ratios for $PM_{2.5}$ and nitrogen dioxide. Indoor $PM_{2.5}$ was found to seldom exceed outdoor levels, except in naturally ventilated buildings, registering an indoor/outdoor ratio close to 1 [Irga and Torpy, 2016]. In the case of nitrogen dioxide, a near constant indoor/outdoor ratio of 1 was recorded, indicating high dependence on environmental conditions [Irga and Torpy, 2016].
Montgomery et al. [2015] found that the sources of VOC and CO_2 are mainly indoor, resulting in indoor/outdoor ratios >1, while $PM_{2.5}$ sources are mainly allocatable outdoors (indoor/outdoor ratio <1).
Challoner and Gill [2014] examined the indoor/outdoor correlations for $PM_{2.5}$ and NO2. While indoor sources dominated $PM_{2.5}$ concentrations in mechanically ventilated buildings, outdoor concentrations were the major factor in naturally ventilated buildings.

Nitrogen dioxide measurements showed that, rather than the concentration at the air intake (on the roof), the street level concentrations dominated indoor levels [Challoner and Gill, 2014]. Thus indicating a major ingress by infiltration.

Influence of equipment Indoor equipment contributes to indoor air pollution by distributing and resuspending, and actively producing pollutants. Cacho et al. [2013] examined office equipment's volatile organic compound and $PM_{2.5}$ emissions. According to them, printers and photocopiers are the dominating indoor sources and emit VOCs and particles during the printing process.
Destaillats et al. [2008] evaluated the pollutant emissions of computers, photocopiers, and printers. The study found that computers source volatile organic compounds and resuspend settled particles by ventilation. Printers emit significantly more volatile organic compounds than computers. Furthermore, they are an active source of $PM_{2.5}$ during their operation [Destaillats et al., 2008]. Potential ozone emissions from photocopiers are discussed in the literature but not finally proven. However, already minimal ozone concentrations together with volatile organic compounds contribute to the generation of harmful secondary pollutants [Destaillats et al., 2008].

Influence of HVAC The HVAC system majorly influences the level of indoor air pollution. Either directly through mechanical ventilation by filtering and conditioning or indirectly through heating and cooling processes.
Irga and Torpy [2016] found a significant correlation between the applied ventilation strategy and indoor pollutant levels. The study was performed on eleven buildings equipped with natural ventilation, mechanical ventilation, and combined natural and mechanical ventilation systems. Overall the lowest pollutant levels were measured in mechanically ventilated buildings, indicated by their significantly lower indoor/outdoor ratio than the other ventilation systems. This is attributable to the filtering of the supply air [Irga and Torpy, 2016]. Natural ventilation and combined ventilation recorded indoor/outdoor ratios close to one due to the unfiltered air supply. Furthermore, higher concentrations and diversity of fungal colonies were recorded in naturally ventilated buildings [Irga and Torpy, 2016]. Maximum carbon dioxide concentrations were discovered in combined ventilation systems of mechanical ventilation and natural ventilation [Irga and Torpy, 2016].
Similar outcomes were reported by Montgomery et al. [2015], with indoor/outdoor ratios close to one in naturally ventilated buildings and significantly lower indoor/outdoor ratios in mechanically ventilated buildings . Furthermore, Montgomery et al. [2015] proved that HVAC systems efficiently controlled volatile organic compound and carbon dioxide concentrations, however, being majorly dependent on the respective ventilation schedules. Carbon dioxide concentrations were reported to be highest in naturally ventilated buildings, exceeding the 1000 ppm threshold at maximum occupancy [Montgomery et al., 2015].
Jiang et al. [2020] ascertained a significant correlation between the recirculation fraction

and the ozone concentration. At increasing recirculation rates (0%, 50%, 100%), the indoor/outdoor ratio of ozone dropped from 0.7 to 0.15 [Jiang et al., 2020], indicating that outdoor air is the predominant source of indoor ozone and that ozone is largely unaffected by standard filters.

Zhang et al. [2020] reviewed the contaminant removal effectiveness and air exchange efficiency of different HVAC system combinations of the cooled ceiling, cooled floor, heated ceiling, heated floor, radiant ceiling, radiant floor, diffuse ceiling ventilation, displacement ventilation, mixing ventilation, personalized ventilation, stratum ventilation, and underfloor ventilation. Displacement ventilation achieves a high contaminant removal effectiveness in the breathing zone due to a stratification of the air, resulting in a zone of cool and fresh air in the lower zone, a transition zone up to 1.7 meters, and an accumulation of warm and polluted air at around 2 meters [Zhang et al., 2020]. However, the contaminant removal effectiveness of displacement ventilation is highly dependent on the temperature differential between the upper and lower zone; thus, if cooled ceilings are employed alongside displacement ventilation impairs air quality since the lower temperature differential leads to increased pollutant concentrations in the breathing zone [Zhang et al., 2020]. The effect applies if displacement ventilation is combined with floor heating systems since increased air temperatures in the lower zone decrease the stratification effect [Zhang et al., 2020]. Conversely, a combination of displacement ventilation and cooled floor systems further increases the temperature differential and improves air quality in the breathing zone [Zhang et al., 2020]. The air quality in mixing ventilation is only weakly related to different heating and cooling systems since it is driven by momentum rather than buoyancy [Zhang et al., 2020]. In most scenarios, pollutants are nearly equally distributed within the zone [Zhang et al., 2020].

Personal ventilation showed the highest removal effectiveness, reducing contaminant concentration up to 84.4 % in the breathing zone, depending on air change rate and source location [Zhang et al., 2020]. On the whole zone level, personal ventilation behaves similarly to mixing ventilation [Zhang et al., 2020].

3.2 Measurement methods

The following section analyzes the state of research regarding measurement methods for in-situ long-term measurements of indoor air pollutants with a high spatiotemporal resolution. This section starts with general classification and delimitation of measurement methods (laboratory measurement methods, in-situ measurement methods). It continues by analyzing the three most common categories of in-situ indoor air pollutant measurement methods (electrical, electrochemical, and optical). The section closes by comparing these aforementioned measurement methods and a assessment of methods that various studies performed.

3.2.1 Classification

In general, measurement methods can be differentiated into laboratory and in-situ measurements. In laboratory measurements, samples are taken on site and transferred for analysis off site to a laboratory. Component and concentration analysis will then be performed using laboratory equipment. Even though laboratory measurements provide the overall highest accuracy, they will not be considered in this review since these are only applicable for singular measurements for one point in time.
In-situ measurements can be further subdivided into low-cost measurement instruments and reference instruments. The latter are instruments defined by their compliance with national and international air quality measurement guidelines and are of high cost from ten to hundred thousands of dollar, high complexity, and stationary [Chojer et al., 2020]. These aspects make the ubiquitous monitoring of indoor air pollutants with reference equipment infeasible and produce insufficient data for calculating spatial pollutant distribution and hotspot identification [Liu et al., 2019]. The main use of reference instruments is the legally required monitoring of air pollutants, in hazardous workplaces or traffic-intensive city areas [Liu et al., 2019]. In the US, the density of reference instrument measurement stations for outdoor pollution is as low as 2 - 5 per 1000 km^2 [Apte et al., 2017]
Low-cost pollutant sensors, on the other hand, are an emerging technology. Costs below 1000 USD per sensor allow the ubiquitous use of pollutant sensors, thus, allowing for a dense sensing network, which enables the identification of pollutant hotspots and distribution. Low-cost pollutant sensors are already applied in portable pollutant monitoring systems [Chojer et al., 2020] and for ventilation control in cars [Frederickson et al., 2021]. Generally, in-situ low-cost pollutant sensors can be differentiated by their measurement principle. The three predominant principles are electrochemical measurements, electrical resistance measurement with metal oxide material (MOS), and optical measurements [Frederickson et al., 2021].
Performance metrics for a sensor are its sensitivity, stability, selectivity, and range of detection. According to Frederickson et al. [2021], sensitivity is "the change of measured signal per change of analyte concentration." The lowest detectable concentration of analyte is the range of detection [Umar and Hahn, 2010]. Stability is "the ability to produce consistent results over a defined period or change in conditions" [Frederickson

et al., 2021] and selectivity states if a sensor responds to a specific target gas or is responsive to a group of gases [Pang et al., 2018]. The measurement principles will be analyzed in the following subsections.

3.2.2 Electrochemical air pollutant measurement methods

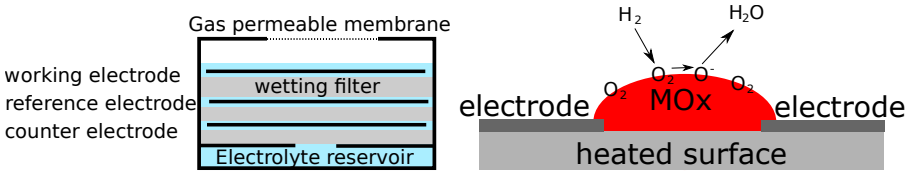


Figure 3.1: Schematic measurement principle of electrochemical (left) and MOS (right) pollutant sensors, own representation based on [Frederickson et al., 2021]

Electrochemical gas sensors use oxidation or reduction reactions between a target gas and a catalyst, inducing a charge between its electrodes [Mead et al., 2013]. Electrochemical sensors can measure concentration in the range of parts per billion to parts per million and are therefore applicable to most indoor air pollutants [Frederickson et al., 2021].
Electrochemical sensors for indoor pollutant measurements mainly apply an amperometric measurement method [Frederickson et al., 2021]. In this, reduction or oxidation reactions between the target gas and a catalysator induce an electric current on the working electrode. The electrolyte guarantees a linear response to increasing concentrations by limiting the uptake by diffusion [Frederickson et al., 2021]. Typically the electrolyte is a mixture of mineral acid and salt [Frederickson et al., 2021]. The working electrode's potential is balanced by the reference electrode and compared to the counter electrode. The current between the working and counter electrode is directly proportional to the concentration of the target gas [Frederickson et al., 2021]. An electrochemical sensor consists of four main parts: the gas permeable membrane, a gas chamber, an electrochemical cell, and an electrolyte reservoir [Frederickson et al., 2021]. For increased specivity, a filter can be added to reduce cross-sensitivities. A typical application for added filters is the measurement of NOx and ozone since these are highly cross-sensitive to each other [Frederickson et al., 2021].
Electrochemical sensors suffer from cross-sensitivities from other pollutants, especially if the concentrations are higher in order of magnitudes. For example, high carbon dioxide concentrations in indoor areas may affect the measurement of pollutants in low concentrations (ozone, nitrogen dioxide) [Frederickson et al., 2021]. In some cases, the cross-sensitive pollutant's response dominates the target gas's response [Pang et al., 2018]. Furthermore, some electrochemical sensors are highly sensitive to changes in relative humidity. In the case of ozone sensors, the correlation towards relative humidity is even higher than towards the target gas [Pang et al., 2018].
The lifetime of electrochemical sensors is limited due to chemical changes in electrolyte

and catalyst [Frederickson et al., 2021]. Periodic replacement or recalibration intervals are therefore necessary in order to upkeep accuracy.

3.2.3 Electrical air pollutant measurement methods

Electrical sensors employ a metal-oxid (MOS) based semiconductor, which changes its resistance depending on the concentration of its target gas. The target gas, to which a specific MOS sensor is sensitive, depends on the MOS compound and its temperature [Frederickson et al., 2021].
The sensing mechanism is based on a change of electrical charge in the MOS-material. The oxidizing target gas acts as an electron acceptor from the MOS material [Frederickson et al., 2021]. Thus, a charge is introduced, leading to the semiconductor resistance change [Frederickson et al., 2021]. The resistance is a function of the charge, dependent on the number of adsorption processes on the MOS surface, which correlates to the target gas concentration. Since the reaction only takes place at above room-level temperatures, a heating element is integrated in order to heat the MOS semiconductor to $200°C$ - $400°C$ [Frederickson et al., 2021]. Some sensors use varying temperatures to target a more specific group of gases by knowing their temperature dependence. The reaction is completely reversible; thus, MOS sensors can theoretically operate for unlimited time [Frederickson et al., 2021].
MOS sensors have a simple structure composed of the MOS semiconductor with two electrodes and a heating element. The resistance measurement is taken between the two electrodes [Frederickson et al., 2021]. To target different gases, electrical sensors use varying metal-oxide compounds. Due to its sensitivity and cost -effectiveness, SnO2 is prevalent and sensitive to a wide array of VOC and oxidizing gases in indoor air. Other common metal-oxides and their respective target gases, as examined in Eranna et al. [2004], are summarized in table 3.4. The major drawback of electrical MOS sensors is

Table 3.4: Common metal-oxide compounds and their target gases

Group	Compound	Target gas
Tin oxides	SnO2	NH3, CH3CHO, CH3SH, H2S, NO2, C2H5OH, CO_2, CO, benzene, o-xylene, SO2, O2, H2, H2O, C4H10, CO, NO2, NOx, petrol, CH4, O3, C3H8
Zinc oxides	ZnO	CO, CH4, H2, H2O, C2H5OH, C3H8, NO, NO2, CO_2, O3, C2H6, n-C4H10
Titanium oxides	TiO2	H2, CO_2, O2, CO, CH4, ethanol, methanol, propanol
Tungsten oxides	WO3	H2S, O3, NO, NO2, NOx, NH3
Indium oxides	In2O3	O3, H2, CO, C3H8, NO2

their low selectivity since various pollutants induce a resistance change in MOS sensors. Furthermore, a significant drift of the resistance values over time through chemical poisoning or degradation of the MOS compound [Frederickson et al., 2021] requires fre-

quent (preferably automatic) recalibration. Due to the selectivity and stability problem, MOS resistance values cannot be directly correlated to the concentration of individual pollutants but are rather suitable for a qualitative assessment of the concentration change of certain groups of gases. This makes MOS sensors especially applicable for measuring volatile organic compounds since their variety makes individual measurements infeasible.

3.2.4 Optical air pollutant measurement methods

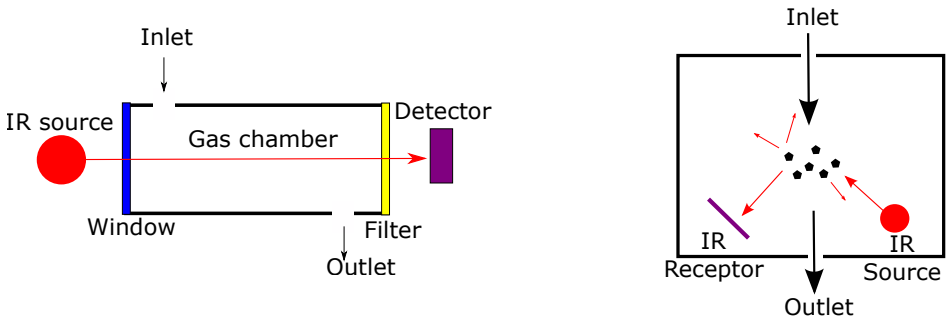


Figure 3.2: Schematic measurement principle of NDIR (left) and laser scattering (right) pollutant sensors, own representation based on [Frederickson et al., 2021]

Optical pollutant measurements employ measurement of the optical characteristics of gases and particles. Light scattering and light adsorption are prevalent in low-cost pollutant sensors. In both cases, a light source emits light of a defined wavelength, passing through a test chamber and measured with a light detector. The amount of detected light indicates the pollutant concentration in the test chamber.
Nondispersive infrared (NDIR) spectroscopy uses the light adsorption characteristics of different gases to detect their concentration. Each gas has a specific infrared wavelength determined by its molecular weight and chemical bonds [Frederickson et al., 2021]. This specific gas absorbs light at this wavelength but is unaffected by other gases. This principle is used in NDIR spectroscopy. Infrared light of the wavelength of the target gas passes through a test chamber with a gas mixture. The light passes through and is partially absorbed by the target gas in the test chamber. At the end of the test chamber, a light detector measures the remaining intensity. The difference in light intensity is correlated to the target gas concentration.
Due to each gas's unique infrared wavelength, NDIR spectroscopy has low cross-sensitivity and high selectivity [Frederickson et al., 2021]. However, it is only applicable to gases at higher concentrations (ppm) and, therefore, cannot detect many pollutants in indoor air. Due to high concentrations in indoor air and high absorptivity in the infrared spectrum, NDIR spectroscopy is the standard for measuring carbon monoxide and carbon dioxide in indoor air [Marinov et al., 2018].
Light scattering sensors or **optical particle counter** (OPC) are common for low-cost monitoring of particle pollutants [Manikonda et al., 2016]. The characteristics of particles

interacting with light are mainly dependent on their size. Particles significantly larger than the wavelength irradiating them tend to absorb the light, and particles significantly smaller have no interaction. Only particle sizes close to the wavelength cause scattering [Frederickson et al., 2021].
In the case of $PM_{2.5}$, the particle sizes are between 1 μm and 10 μm; therefore, OPC sensors commonly employ infrared light sources to detect particles [Frederickson et al., 2021]. Due to the choice of light source, particles beneath 1 μm and above 10 μm are hardly detected.
Structurally OPCs are similar to NDIR sensors. The gas mixture is injected into a test chamber and irradiated by an infrared light source. Particles in the gas mixture cause a scattering of infrared light, and some scattered light hits a light detector. The amount of scattered light, and thus light intensity at the detector, stands in relation to the particle concentration in the gas mixture [Frederickson et al., 2021].
However, other pollutant characteristics such as refractive index and light absorptivity also influence the amount of scattered light, and therefore to a potential misinterpretation of the results [Frederickson et al., 2021]. Furthermore, OPCs were shown to have a size bias, overestimating certain particle sizes while underestimating others. Additionally, environmental conditions such as high humidity lead to a significant overestimation of $PM_{2.5}$ in OPCs due to the scattering and adsorption of light in water droplets [Frederickson et al., 2021]. Therefore, most OPCs are designed for an operating range of up to 80% relative humidity.

3.2.5 Field assessment of in-situ pollutant sensors

Various studies assessed the potential of low-cost in-situ pollutant monitors in the field compared to reference instruments.
Demanega et al. [2021] examined the performance of low-cost monitors in a test chamber setup and compared them to reference instruments. The test was performed under varying pollutant events, such as a burning candle, vacuuming or use of room deodorant, and different indoor climatic conditions. Demanega et al. [2021] assessed the measurement of volatile organic compounds, fine and coarse particulate matter, and carbon dioxide. Seven carbon dioxide sensors were assessed and compared to a reference instrument. Six of the seven sensors applied the NDIR measurement method; one derived it from a VOC sensor (eCO2). All NDIR sensors performed well compared to the reference unit; the Pearson correlation coefficient exceeded 0.97 for all but one device, which had a delayed response and thus only reached 0.8 [Demanega et al., 2021]. The eCO2 sensor did not correlate with the reference instrument and had a negative Pearson correlation coefficient of -0.36. The Sensirion SCD40 NDIR sensor performed best with a Pearson correlation coefficient of 0.99, a mean relative error of 6%, and an error of 3% for the reported peak concentration [Demanega et al., 2021].
Ten particulate matter sensors were compared to a reference instrument for fine and coarse particles. All sensors used the OPC measurement method. The tested pollutant scenarios generated a wide range of particles ranging from 0.2 μm to 13.1 μm. The low-

cost $PM_{2.5}$ sensors mostly underreported $PM_{2.5}$ concentration; however, most achieved a strong correlation towards the reference instrument [Demanega et al., 2021]. Many sensors performed differently for the varying pollutant sources, showing a size bias. The Sensirion SPS30 performed most stable with a Pearson correlation coefficient of over 0.737 in the worst scenario and 0.998 in the best; however, it significantly underreported most concentrations by up to 73%. Three sensors were examined regarding their VOC measuring capability. All of these applied the MOS measurement method . All sensors showed a high correlation to reference instruments. However, the absolute values highly varied, with some sensors over and others underreporting. The best results were achieved by the Sensirion SGP30 sensor with a mean relative error of 17% and a strong Pearson correlation coefficient [Demanega et al., 2021].
Liu et al. [2019] examines low-cost $PM_{2.5}$ sensors, employing the OPC measurement method, in outdoor air and compares them to reference air quality monitoring stations in Oslo. Liu et al. [2019] located three sensors next to the reference station and evaluated their results regarding the accuracy, response, inter-sensor variability, and environmental impact factors. Over four months, measurements showed a strong correlation for two of the three sensors with R^2 values of 0.71 and 0.68 and one with slightly less accuracy at 0.55. The long-term average showed high accuracy above 80% and up to 98% for one sensor. In comparison to each other, the low-cost sensors show little variability, around 10%. The values of the individual sensors correlate very strongly with R^2 above 0.97 for all sensors [Liu et al., 2019]. Furthermore, Liu et al. [2019] found a significantly diminished accuracy for environmental conditions with relative humidity above 80%.
Badura et al. [2018] examined the performance of low-cost fine $PM_{2.5}$ sensors towards reference air quality monitoring stations in Wroclaw. Four types of OPC sensors with three identical units each were collocated to the reference station and monitored for nearly six months [Badura et al., 2018]. The results indicate a strong correlation with R^2 values above 0.74 for all units. All units showed a continuous output over the measurement period without significant shifts in the measurement results [Badura et al., 2018]. Relative humidity above 80% showed to majorly impair the results [Badura et al., 2018].
Borrego et al. [2016] assessed the performance of electrochemical and MOS sensors, measuring nitrogen dioxide and ozone, within the EuNetAir program. Of the sensors, which applied an electrochemical measurement method, the majority achieved a high Pearson correlation coefficient above 0.89 for nitrogen dioxide and mixed results for ozone, with individual sensors achieving 0.88. In contrast, others showed next to no correlation [Borrego et al., 2016]. In the MOS sensors, none could achieve a correlation for nitrogen dioxide, and only one correlated with ozone.

3.3 Virtual sensing of indoor air pollutants

To achieve widespread monitoring of indoor air pollutants, it's essential to employ sensors that are not only cost-effective and energy-efficient but also robust and require minimal maintenance. Technologies like NDIR and OPC, which rely on optical measurement techniques, often involve delicate components. These technologies face challenges such as measurement drift and reduced lifespan [Kolarik et al., 2020], largely due to the accumulation of external particles in the measuring areas. Additionally, these optical methods tend to consume more energy, making them less ideal for battery-powered applications. On the other hand, MOS-based VOC sensors, while being economical and energy-efficient, are susceptible to drift and often encounter issues with consistent performance [Frederickson et al., 2021]. Consequently, there's a need to explore alternative methods for monitoring particulate matter ($PM_{2.5}$), volatile organic compounds (VOC), and carbon dioxide (CO_2).
Utilizing virtual sensing for $PM_{2.5}$, VOC, and CO_2 presents an alternative to the widespread deployment of physical sensors. Virtual sensing "aims to approximate unmeasured physical quantities in a dynamic system using existing sensor information. This is especially beneficial when important locations of the system are difficult to instrument, or the cost of sensors is very high" [Heindel et al., 2021]. "A virtual sensor uses low-cost measurements and mathematical models to estimate a difficult to measure or expensive quantity" [Li et al., 2011]. These models utilize associated physical measurements, control signals, operational data, and design details [Yoon, 2022]. The use of virtual sensing is extensively seen in various fields like process control, automotive, aviation, and robotics [Li et al., 2011]. Nevertheless, the rapid increase in available data points, propelled by advancements in IoT and the decreasing cost of sensors, has made virtual sensing an increasingly researched topic in the building industry [Yoon, 2022]. In the context of the building industry, the applications of virtual sensing are diverse. Buildings produce numerous data points, and almost every physical sensor can contribute extra information for virtual sensing purposes. Li et al. [2011] provide an example of virtual sensing's potential in buildings: "A 'smart' lighting fixture could provide power, lighting, and heat gain outputs based on the input control signal. A 'smart' window could provide estimates of heat gain and even solar radiation based on low-cost measurements and a model."
In building environments, virtual sensing applications encompass monitoring HVAC operations [Wu et al., 2016; Hong et al., 2021], calculating indoor infiltration rates [Li et al., 2019], determining zone temperature distributions [Alhashme and Ashgriz, 2016], estimating zone occupancy [Zhao et al., 2015], and monitoring indoor air pollutants [Skön et al., 2012; Elbayoumi et al., 2015; Khazaei et al., 2019].

3.3.1 Machine learning in virtual sensing

Artificial intelligence (AI) is a broad term encompassing various technologies that enable machines to mimic human intelligence. Initially, AI relied on rule and knowledge-based systems as primary methods for problem-solving. These systems necessitated the explicit programming of rules by experts to replicate human decision-making processes. However,

their ability to tackle complex or unfamiliar problems was limited, and they lacked the capacity to learn from experience [Russell and Norvig, 2010]. Subsequently, the evolution of AI introduced probabilistic methods to address issues beyond the scope of rule-based systems. Unlike their predecessors, probabilistic methods do not require the explicit programming of rules, a paradigm commonly known as machine learning (ML). Machine learning involves training computers to learn from data derived from past experiences [Djenouri et al., 2019], proving useful in scenarios where intricate or unclear relationships between input and output data preclude the use of rule-based systems. Applications of ML span various fields, including image and speech recognition, and natural language processing. ML systems are typically categorized as black-box models, which are systems whose internal mechanisms are opaque and cannot be understood by examining their parameters, such as neural networks [Molnar, 2020]. In contrast, rule-based systems are considered white-box models, characterized by their transparent internal mechanisms. A grey-box model represents an intermediate, with partially known mechanisms. Recent advancements in explainable AI (XAI) focus on transforming black-box models into grey-box models by making their internal mechanisms more transparent and understandable [Molnar, 2020]. Explainability is crucial in domains requiring comprehension of the model's decision-making processes, such as in medical diagnostics and autonomous vehicles [Molnar, 2020].

ML is categorized into three primary approaches based on the nature of the available data: supervised, semi-supervised, and unsupervised learning. Supervised learning necessitates a substantial amount of training data with known output values, often requiring manual labeling or annotation [Russell and Norvig, 2010]. When labeling extensive datasets is impractical or cost-prohibitive, semi-supervised learning is employed, leveraging a small set of labeled data combined with a larger set of unlabeled data to train the model. Conversely, unsupervised learning operates without ground truth, aiming to discern patterns and regularities within the training data that can then be applied to detect similar patterns in new datasets [Russell and Norvig, 2010].

Ground truth for supervised or semi-supervised learning can be generated by measuring the target value or manually labeling the data. In the training procedure, the ML model attempts to identify regularities in the training data; the identified connections can then be used to detect similar regularities in new data [Djenouri et al., 2019].

For virtual sensing of indoor air pollutants supervised ML is applicable since the ground truth can be measured for indoor pollutants using previously mentioned measurement equipment. To train the model, extensive data is required. In the building stock, the necessary data is available in BMS systems. BMS data has become standard in modern construction and is progressively being integrated into older structures via the installation of wireless BUS technologies like LoRaWAN. These systems collect a wide range of data, including information from actuators, water, air, and electrical systems, internal environmental conditions, and the activities of building occupants.

ML problems can be further differentiated into classification and regression problems. In classification, the machine learning algorithm decides which discrete class an observation belongs to [Qolomany et al., 2019]. Regression problems require a continuous numerical output and usually involve finding a mathematical function that maps the input variables

to the output data. A common supervised ML method are artificial neural networks (ANN). ANNs imitate the working principle of neurons in the human brain. Several layers of neurons are connected with variable weights and triggered by activation functions [Qolomany et al., 2019]. A special case of ANNs are long-short-term memory recurrent neural networks (LSTM). Studies in various fields have demonstrated the effectiveness of LSTMs for managing time-series data in the development of virtual sensors. This is attributed to their design, which allows for the integration of historical data into their predictive models. The unique structure of LSTMs, which includes memory cells in their network design, facilitates the recognition and retention of temporal patterns in data series [Heindel et al., 2021]. In the realm of building management, the application of LSTMs is evident in areas such as demand side management by forecasting energy usage [Karijadi and Chou, 2022; Jang et al., 2022] and in estimating building occupancy [Qolomany et al., 2017]. However, the use of LSTMs in simulating virtual sensors for indoor air pollution has not been explored yet.

The ML process can be summarized in five steps: data acquisition, preprocessing,

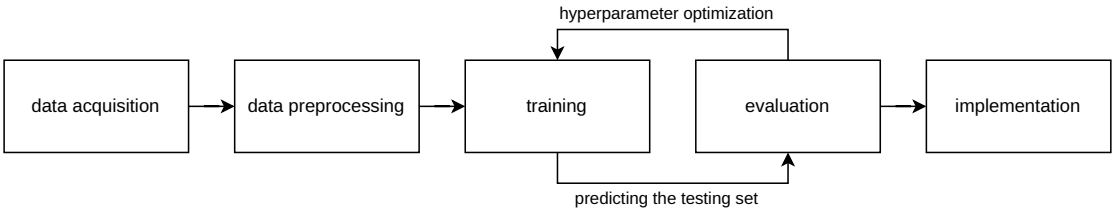


Figure 3.3: Schematic machine learning process

learning, evaluation and implementation. Data acquisition includes all tasks necessary to gather input and ground truth data through measurements, surveys, archive data, and open datasets [Djenouri et al., 2019]. Data preprocessing involves all steps necessary to process the raw data to increase learning performance. This includes data enrichment by adding contextual data such as date, time, and location to the raw data set and cleaning the raw data by removing and eradicating measurement errors. Next is selecting the prediction features by performing a correlation analysis, filtering out data with low or no correlation towards the output variable, and reducing the dimensionality of the data set by combining features. The last preprocessing step involves splitting the data into a training and testing set to evaluate prediction performance later. In the learning phase, an ML algorithm is used to identify relationships between the input and output data of the test data set. Depending on the method and size of the data set, learning can be resource and time intensive. Evaluation interprets the performance of the trained ML model on the unknown test data set. The known truth value is compared to the predicted output of the ML model and evaluated using prediction accuracy metrics. Implementation is the last step and involves deploying the trained and evaluated model to its designated task.

Various studies previously explored the application of ML models based on BMS data in the following application scenarios: prediction of occupancy and activity recognition [Qolomany et al., 2019], thermal comfort estimation [Djenouri et al., 2019], energy demand

prediction [Djenouri et al., 2019], and optimization of energy efficiency by device control [Qolomany et al., 2019]. Further applications include: activity estimation, behavior recognition, fault detection, and sensor metadata inference [Djenouri et al., 2019].

3.3.2 Differentiation of virtual sensors

Virtual sensors are primarily categorized into three functional areas: replacement, observation, and assistance, as outlined in [Yoon, 2022]. Sensors of the replacement type function alongside their physical counterparts. They are capable of identifying issues like sensor malfunctions or calibration discrepancies by analyzing the difference between physical and virtual sensor readings [Kusiak et al., 2010]. Moreover, these sensors can substitute their physical equivalents when necessary [Yoon, 2022]. Observation-type virtual sensors calculate certain data points in the absence of physical sensors, utilizing alternative measurements and mathematical models [Yoon, 2022]. Conversely, assistance virtual sensors, rather than measuring a physical parameter, are used in conjunction with other virtual sensors to enhance precision, often producing outputs that are standardized [Yoon, 2022]. Moreover, virtual sensors can be distinguished based on their modeling approach and the nature of the measurements they rely on [Li et al., 2011]. The modeling approaches include white-box, grey-box, and black-box models [Li et al., 2011], while the measurements are categorized into transient types (such as power consumption, indoor temperature) and steady-state types (like system malfunction status) [Li et al., 2011]. White and grey-box models necessitate comprehensive understanding of the building, making them impractical for older buildings due to the lack of accessible building information, unrecorded modifications, and degradation in performance. An alternative approach is the black-box model, which demands extensive data from measurements. In conclusion, for predicting indoor air pollutants in the non-residential building stock, transient-state observational virtual sensors, developed through black-box modeling techniques, are essential.

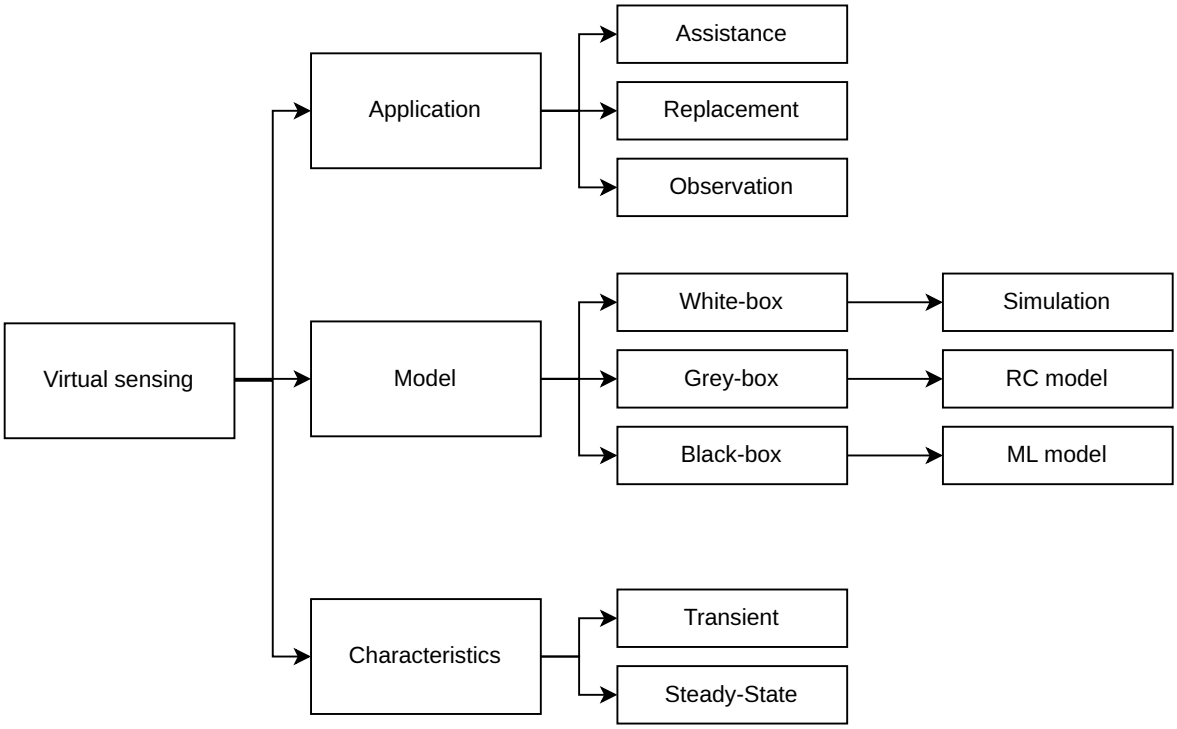


Figure 3.4: Differentiation of Virtual sensors own representation based on [Li et al., 2011]

3.3.3 Virtual sensing in indoor air pollutant prediction

Prior research has explored the potential of virtual sensing in predicting indoor air pollutants. Kusiak et al. [2010] developed replacement virtual sensors for measuring temperature, humidity, and CO_2 utilizing four different modeling techniques to calibrate and supervise physical sensors, integrating data from HVAC systems, climate conditions, and other indoor air pollutants. It was determined that MLPs (Multi-Layer Perceptrons) were most effective in simulating these physical sensors. Kusiak et al. [2010] concluded that virtual sensors are capable of identifying malfunctions in their physical counterparts and substituting them when needed. Skön et al. [2012] implemented MLPs to construct an observation virtual sensor, taking temperature and humidity data to model CO_2 levels. Skön et al. [2012] found that predicting CO_2 levels based solely on temperature and humidity is challenging, necessitating additional data to enhance the accuracy of the black-box model. Leidinger et al. [2014] engineered a replacement virtual sensor for selective detection of VOCs such as formaldehyde, benzene, and naphthalene, using a variety of low-cost MOS sensors as inputs. In their approach, Leidinger et al. [2014] employed linear discriminant analysis for estimating these specific substances. While the study achieved a high classification accuracy of over 99% in laboratory settings, this accuracy significantly declined to 83% in field tests, attributed to the VOC emissions from the hardware itself.

Table 3.5: Summary of Literature on Virtual Sensor Creation.

Study	Output	Methods	Main Findings
[Kusiak et al., 2010]	Temp., Hum., CO_2	MLP, SVM, Pacereg, RBF	MLP outperformed other models; Virtual sensors can detect and replace failing physical sensors
[Skön et al., 2012]	CO_2	MLP	Estimating CO_2 based only on temperature and humidity is challenging
[Leidinger et al., 2014]	VOC	Linear Discriminant Analysis	99% lab accuracy, 83% field accuracy due to hardware VOC emissions
[Karijadi and Chou, 2022] [Jang et al., 2022]	Energy	LSTM	LSTMs have been successfully applied in energy consumption forecasting
[Qolomany et al., 2017]	Occupancy	LSTM	LSTM can be used for predicting occupancy

3.4 Demand controlled ventilation based on indoor air pollutants

Above sections focused on acquiring indoor air pollutant data either via physical sensors or virtual sensing. Mitigation, however, requires active implementation measures. Tham [2016] enumerates the following approaches to improve the health of occupants by employing pollutant monitoring:

- **Exposure monitoring**: assessing the individual pollutant exposure and accordingly reduce exposure time in polluted areas
- **Pollutant removal**: remove pollutants from the air using air filtering approaches.
- **Demand controlled ventilation**: Pollutant-dependent outdoor/clean air flow in order to dilute pollutants

Pollutant removal and demand-controlled ventilation actively contribute to removing and diluting pollutants in indoor air. However, Zhang et al. [2011] notes that no air-cleaning technology is effective against the whole pollutant spectrum. Luengas et al. [2015] came to the same conclusion and ascertained that no technology is suitable for generating satisfactory indoor air quality without further measures. Thus, in the current state of technological development, pollutant dilution is the only active measure effective against the whole pollutant spectrum.
Technologically pollutant dilution is implemented by demand-controlled or personal

ventilation. Demand controlled ventilation "denotes continuously and automatically adjusting the ventilation rate in response to the indoor pollutant load" [Mansson et al., 1997]. It is an additional, sensor data-based feedback layer to scheduled ventilation control [Coleman and Meggers, 2018], which typically utilizes humidity, carbon dioxide, or motion detectors as a proxy for indoor air quality [Coleman and Meggers, 2018].

3.4.1 Control strategies

The main goals of demand-controlled ventilation are to eradicate undersupply during occupancy and oversupply during unoccupied times, thus reducing air treatment and transportation energy consumption and improving indoor air quality. Furthermore, infection prevention became an important task of demand-controlled ventilation during the Covid pandemic.
This means that demand-controlled ventilation "operates at reduced airflow rates during a large amount of the operation time and thus consumes less energy for fan operation and heating/cooling the supply air" [Merema et al., 2018]. Demand-controlled ventilation is implemented through variable air volume systems, which control the supply of air volume to a zone via controllable valves [Anand et al., 2019b] in a centralized HVAC system or via individual control of decentralized units.
Different control strategies can be based on occupancy estimation, pollutant, and thermal comfort data. Therefore, optimizing the minimum required ventilation rate based on occupancy, controlling the temperature, and controlling the pollutant concentration are the levers [Anand et al., 2019b].

Minimize oversupply The least invasive strategy, as described by Anand et al. [2019b], is the sole control of the occupant based required minimum ventilation rate according to the occupancy estimation. The thermal comfort and pollutant dilution are upkept during unoccupied hours [Anand et al., 2019b].

Minimize oversupply and thermal conditioning Anand et al. [2019b] examined a strategy to expand the comfort band during unoccupied hours to reduce heating and cooling energy. In order to avoid discomfort upon reentry into the space, Anand et al. [2019b] suggest setting an upper limit of 28 $°C$.

Restrict conditioning to occupancy The most invasive control strategy suggested by Anand et al. [2019b] is to completely discontinue conditioning of zones unoccupied for more than one day in combination with the previous measures.
In the built environment, demand-controlled ventilation is usually built around estimating occupancy in order to supply only the necessary amount of air during occupied hours [Anand et al., 2019b]. In some cases, this is complemented by CO_2 or VOC measurements in the exhaust air duct [Coleman and Meggers, 2018] and control strategies to improve thermal comfort [Anand et al., 2019b].

3.4.2 Occupancy proxies in demand controlled ventilation

Since occupancy is not directly measured in the building stock, different proxies are applied to estimate it. Four methods are mainly employed in order to estimate the number of occupants: humidity-based occupancy estimation [Pecceu et al., 2018], carbon dioxide-based occupancy estimation [Pantazaras et al., 2018], motion-based occupancy estimation [Coleman and Meggers, 2018] and camera-based occupancy estimation [Anand et al., 2019b]:

Humidity-based Pecceu et al. [2018] investigated the applicability of humidity-based-occupancy estimation and conducted annual measurements in residential and nonresidential buildings. The study assumed carbon dioxide concentration as the ground truth for occupancy and examined the correlation between it and relative humidity [Pecceu et al., 2018]. The results indicated that relative humidity is not correlated to occupancy, a higher, but still weak correlation was identified for the absolute moisture content in the air [Pecceu et al., 2018]. A moderate correlation could be identified for occupancy and indoor outdoor absolute moisture differences, and the derivatives of carbon dioxide and absolute moisture [Pecceu et al., 2018].

Carbon dioxide-based According to Pantazaras et al. [2018], CO_2 is a good proxy for human occupancy since indoor CO_2 is mainly generated by human respiration. Mass balance equations allow a good estimation of generated CO_2 in a room and thus estimation of occupancy [Anand et al., 2019b]. However, this requires a contained zone with known inflows and outflows. Natural ventilation or infiltration decreases the accuracy of CO_2 as occupant proxy [Anand et al., 2019b]. In the case of a fully contained zone, CO_2 measurements achieve a reasonable prediction accuracy for high occupant density and low prediction accuracy for low occupant density [Pantazaras et al., 2018].

Camera-based Camera-based systems use facial recognition techniques to count the number of occupants in a room [Anand et al., 2019b]. Anand et al. [2019a] and Yang et al. [2018] proved a accuracy of 95% using facial recognition for occupancy detection. However, privacy and data protection concerns make a camera-based occupancy detection system infeasible in nonresidential typologies.

Motion based Motion detectors use passive infrared sensor modules and usually implement a timer reset every time motion is detected [Coleman and Meggers, 2018]. However, Coleman and Meggers [2018] noted that motion detectors are problematic for occupancy estimation in many situations since unmoving occupants are not detected, and passersby distort results.

3.4.3 Robustnes of demand controlled ventilation systems

The controller, sensors, and actuators required for demand-controlled ventilation introduce further points of failure into the building system. Therefore, it is vital to evaluate the robustness of the individual components and identify the possible points of failure. Studies show that problems regarding demand-controlled ventilation are common, especially concerning the placement and failure of sensors used for control.

Sensor placement Studies examining the effectiveness of demand-controlled ventilation in the building stock found that the placement in the sensor is critical for achieving good control. Several studies examined the sensibility of demand-controlled ventilation systems to the placement of the sensor, which could, in the worst case, lead to an undersupply of fresh air due to underestimation of the pollutant concentration or an oversupply if the sensor is located close to pollutant sources.
Pantazaras et al. [2018] examined the placement of CO_2 sensors in a lecture hall, which acts as a proxy for occupation and controls the supply air volume of mixing ventilation. The study concluded that the distribution of CO_2 in the lecture hall was nearly uniform with maximum deviations of 220 ppm; however, most were below 60 ppm [Pantazaras et al., 2018]. Furthermore, no significant stratification effect at different heights could be identified. Thus Pantazaras et al. [2018] summarized that the effect of sensor placement on demand-controlled ventilation is negligible in this case.
Pei et al. [2019] examined the effect of measuring CO_2 at different heights in a displacement and mixing ventilation to identify the magnitude of stratification. Sensors were placed at 0.3 meters, 1.2 meters, and 2.4 meters and resulted in a CO_2 differential up to 800 ppm at 1.2 meters and 900 ppm at 2.4 meters in the displacement ventilation scenario at a fixed air change rate of 2.5 1/h [Pei et al., 2019]. A doubling of the air change rate led to a halving of the CO_2 differential [Pei et al., 2019], showing that air change rate and stratification are indirectly proportional to each other in displacement ventilation. In the case of mixing ventilation, no significant stratification could be discovered, thus implying that the placement of sensors is crucial in the case of displacement ventilation and negligible in mixing ventilation scenarios [Pei et al., 2019]. [Pei et al., 2019] advises placing sensors in displacement ventilation in the breathing zone at about 1.0 - 1.2 meters. However, removed from occupants in order to avoid measuring increased concentrations in the thermal plume [Pei et al., 2019].

Sensor failure Sensors are highly sensitive electronic equipment and are prone to fail if not handled correctly or after a certain time of application. NDIR and OPC technology are optical measurement principles that require fragile components and suffer from measurement drift and longevity issues [Kolarik et al., 2020] due to the build-up of foreign particles in the measurement chambers. Furthermore, MOS-based VOC sensors are also prone to significant drift of the resistance values over time through chemical poisoning or degradation of the MOS compound [Frederickson et al., 2021]. Therefore, frequent maintenance and recalibration of sensor equipment is key in ensuring the proper operation of a demand-controlled ventilation system. Approaches using virtual sensors deployed in

parallel to their physical counterparts seek to remedy this problem by detecting sensor faults, and calibration drifts [Kusiak et al., 2010] and even replacing the physical sensor automatically if needed [Yoon, 2022].

3.4.4 Impact of demand controlled ventilation

Demand-controlled ventilation has been shown to significantly impact energy consumption while upholding or improving indoor air quality and thermal comfort. According to Guyot et al. [2018], "energy savings up to 60% can be obtained without compromising IAQ, even sometimes improving it".
In the following, studies are reviewed regarding the impact of demand-controlled ventilation on energy consumption, indoor air quality, and thermal comfort.

Energy consumption reduction Merema et al. [2018] examined the gains of demand-controlled ventilation compared to constant ventilation in three nonresidential buildings, including a school Classroom, a lecture hall, and an office in Belgium. Measurements have been performed during the heating period for at least two weeks [Merema et al., 2018]. Results show that fan energy could be reduced by at least 50% for all buildings and ventilation heat losses reduced by 34 - 47 %. The highest savings were achievable in the lecture hall, reducing fan energy by 55% and heat losses by 47%, attributable to the varying occupancy [Merema et al., 2018]. The energy reduction potential was compared to a constant airflow rate at design occupancy [Merema et al., 2018].
Sun et al. [2011] examined the energy reduction potential of demand-controlled ventilation with a carbon dioxide-based occupancy estimation system in Hong Kong, identifying a 52% reduction potential in fan energy consumption.
Stein et al. [2007] measured the performance of demand-controlled ventilation systems in office buildings in California. Stein et al. [2007] ascertains a reduction potential by 51 - 62 % for fan energy and 34% - 48% for ventilation heat losses.
Mysen et al. [2005] compared two demand-controlled ventilation strategies (CO_2 based, motion-based) to constant air ventilation in 81 schools in Norway. The study concluded that CO_2-based demand-controlled ventilation could achieve a 38% reduction of total HVAC energy and motion-based demand-controlled ventilation 51% [Mysen et al., 2005]. The reduction potential was attributed to the Classroom usage patterns, which were in mean only in use during 30 - 50% of the HVAC operation time [Mysen et al., 2005].
Wachenfeldt et al. [2007] compared a CO_2-based demand-controlled ventilation strategy to a constant ventilation system in two Norwegian schools and found a fan energy saving potential of 87%, while reducing ventilation heat losses by 21%.
Ahmed et al. [2015] examined occupancy-based demand-controlled ventilation in a Finish office building for different control and equipment scenarios. By implementing demand-controlled ventilation over constant ventilation, energy savings of 33 - 41% could be achieved [Ahmed et al., 2015].

Indoor air quality improvement Besides reducing the HVAC energy, upholding or improving indoor air quality becomes a central goal. As stated by Afroz et al. [2020],

"modulating the ventilation rate based on CO_2 sensor data provides better control over IAQ compared with an occupancy time schedule" and thus allows for improved indoor air quality at reduced energy consumption.
Merema et al. [2018] assessed the indoor air quality in four nonresidential buildings with demand -controlled ventilation in Belgium with distributed CO_2 measurements. Merema et al. [2018] found that the measurements complied with the regulative threshold of 1000 ppm most of the time, with brief overruns during proximity of the sensors to occupants [Merema et al., 2018]. Furthermore, the ventilation efficiency (ratio of extract zone to breathing zone concentration) was analyzed and performed well (above 0.89) for all buildings [Merema et al., 2018].
Ahmed et al. [2015] compared the indoor air quality of a Finish office building with demand-controlled ventilation to a constant air volume system for different technological and control scenarios and found that, even though some scenarios lead to worse indoor air quality in demand-controlled ventilation, most achieved equal air quality.
Anand et al. [2019b] examined the effect of demand-controlled ventilation on indoor air quality in office, seminar, and lecture rooms at the University of Singapore and compared it to a constant ventilation system. The results showed that the constant ventilation system significantly oversupplies the rooms at most times while undersupplying seminar rooms with high occupancy in the evening hours. In comparison, demand-controlled ventilation always performed above the required minimum ventilation rate, reducing the exceedance of the constant air volume system [Anand et al., 2019b]. The increased air demand during evening hours is met in the case of demand-controlled ventilation, thus improving the indoor air quality [Anand et al., 2019b].

Infection prevention The covid-pandemic had a major impact on how people use and interact in nonresidential buildings and the requirements for HVAC systems since indoor locations are predominant in infection spread. Schoen [2020] published measures to reduce the infection risk, including an increased ventilation rate, eliminating recirculation, and disabling demand-controlled ventilation. Furthermore, [Schoen, 2020] advised the improvement of the filtration system and continuous operation of the ventilation and mobile air purifiers. Since then, research has been done regarding the compatibility of infection prevention and energy reduction to be prepared for future pandemics.
Wang et al. [2021] implemented an algorithm for demand-controlled ventilation during a pandemic with camera-based occupancy estimation for public buildings. The camera-based system uses machine learning methods in order to detect occupant densities. If the system recognizes distances between persons below 2 meters, an anti-infection mode is activated, which supplies increased air volume to the specific area [Wang et al., 2021]. Otherwise, the ventilation system operates in regular demand-controlled ventilation mode. In a test case, the ventilation system achieved an 11.7% energy consumption reduction compared to a constant volume system while reducing infection probability below 2% according to the Wells-Riley model [Wang et al., 2021].
Pang et al. developed an algorithm for demand-controlled ventilation based on an artifi-

cial neural network assessing infection risk. The artificial neural network was trained against CFD simulations and uses geometry, CO_2 concentration, and outdoor air flow rate as input in order to estimate infection risk [Pang et al.]. The study ascertains that implementing the control algorithm energy efficiency and infection prevention can be achieved [Pang et al.].

3.4.5 Controllers in demand controlled ventilation

Controllers seek to achieve a control goal, either a fixed setpoint or a variable range (tracking control) [Afram and Janabi-Sharifi, 2014]. In general, controllers for demand-controlled ventilation can be differentiated into classical controllers, hard controllers, soft controllers, and hybrid controllers [Afram and Janabi-Sharifi, 2014].

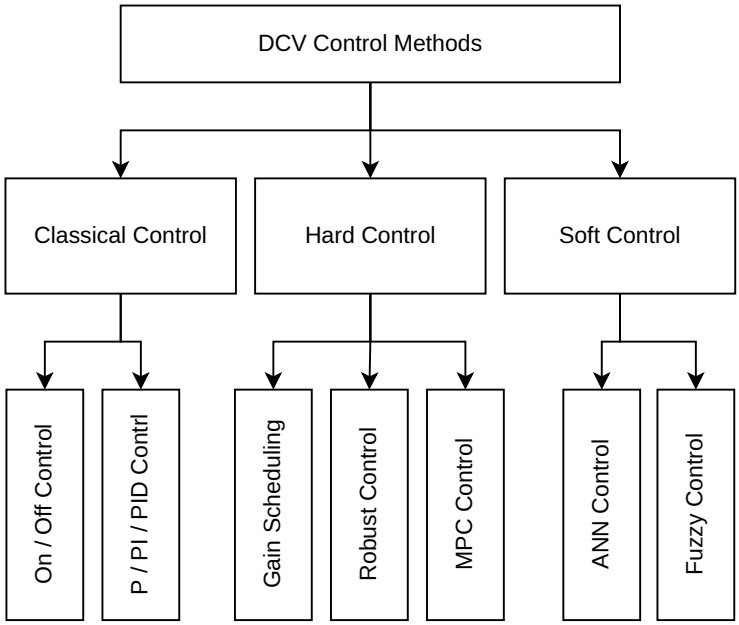


Figure 3.5: Overview of DCV control, own representation based on [Afram and Janabi-Sharifi, 2014]

Classical controllers are the most common control technique in demand-controlled ventilation. Classical control uses On/Off or P, PI, or PID control. The On/Off hystheresis control typically operates within a defined upper and lower band and outputs a binary signal [Afram and Janabi-Sharifi, 2014]. However, the On/Off hystheresis tends to perform unreliably with time-delayed processes. The P, PI, or PID controller output a continuous value within a specified range and works on proportional integration principles to minimize the difference between supply and demand [Lauckner and Krimmling,

2020] . In the case of demand-controlled ventilation systems, a PI controller is typically used to regulate the supply air volume; the proportional part of the controller defines the ventilation demand based on the difference between desired and actual ventilation performance, and the integral part of the controller minimizes the influence of slowly changing disturbances on the control of ventilation performance [Lauckner and Krimmling, 2020]. Even though classical controllers are the most widespread type of controllers in demand-controlled ventilation [Lauckner and Krimmling, 2020], setting the thresholds for On/Off hystheresis and tuning the parameters of PI controllers is often challenging and systems tend to perform unreliable outside of the tuning band [Afram and Janabi-Sharifi, 2014].

Hard controllers are control techniques based on simplified mathematical models of the actual system. Creating hard controllers requires extensive analysis and system state observation. Common types of hard controllers include gain scheduling control, robust control, and model predictive control. Gain scheduling control uses a set of different controllers to react to different operating conditions in cases where a nonlinearity makes it infeasible to use a single controller [Afram and Janabi-Sharifi, 2014]. Robust control is designed to handle errors, uncertainties, and variations in the controller input without major disturbances to the output signal. Its goal is to achieve a stable control that performs well under a wide range of different scenarios [Afram and Janabi-Sharifi, 2014]. Model predictive control uses simplified mathematical models of the physical processes to predict future behavior [Scheuring et al., 2022]. Based on this prediction, an output control signal is calculated by optimizing a cost function for the prediction window [Afram and Janabi-Sharifi, 2014]. The predictions of the model predictive control are updated at each timestep by the actual values.

Soft controllers are control techniques that are based on heuristical models [Afram and Janabi-Sharifi, 2014]. Soft controllers do not require mathematical analysis or tuning like classical or hard controllers. However, soft controllers require extensive system data to achieve good control results. Examples of soft controllers are artificial neural networks or fuzzy controllers. Artificial neural networks are self-learning black box models that use training data to fit a nonlinear algorithm. Fuzzy control is based on linguistic rule descriptions using if, then, and else relationships. Input values are categorized using linguistic terms to arrange them in fuzzy sets (e.g., cold, moderately warm, warm, hot) [Lauckner and Krimmling, 2020]. Rules are then defined based on these fuzzy sets (e.g., if the temperature is hot, the ventilation is high).

Hybrid controllers are a mixture of these controller types. Soft controllers often supplement hard controllers to operate under a wider range of conditions.

3.5 Summary and Contribution

In the following section, the findings from the literature review are summarized, the identified gaps in the literature discussed, and the aspired contribution of this dissertation stated.

3.5.1 Review Conclusion

To summarize the findings of the literature review, this study looked at the state of research in four fields: indoor air pollutants in nonresidential buildings, the measurement of indoor air pollutants using physical sensors, the replacement of physical sensors in indoor air pollutant monitoring using virtual sensing and the application of indoor air pollutant monitoring data in demand-controlled ventilation.
In the first part, the state of research monitoring indoor air pollutants with physical sensors was examined by looking at 15 studies in which 67 nonresidential buildings and 194 rooms were examined, with 62 commercial and eight educational buildings. Most of the reviewed studies (70% where the information was given) measured indoor air pollutant concentration in open office spaces, and each study assessed at least one building with a full mechanical ventilation system. Only in six studies indoor air pollutant concentration was compared to mechanically ventilated buildings with naturally ventilated ones. Notably, half of the reviewed studies only performed measurements for a week or less, and only two studies performed measurements for a full year.
In the literature review, the spatiotemporal resolution of the measurements was assessed. The reviewed studies chose a median sampling of 5 minutes, even though multiple sources identified a high spatiotemporal variation of indoor air pollutants requiring a higher sampling interval. Only one reviewed study examined the spatial distribution of pollutants within a room by deploying multiple sensors.
The most measured pollutant in all studies was $PM_{2.5}$ (14 out of 15 studies), followed by volatile organic compounds (8/15 studies), CO_2 (7/15 studies), CO and NO2 (5/15 studies each), and ozone and SO_2 (2/15 studies each). The studies also evaluated the relevance of pollutants regarding occupant health impact. It was found that carbon monoxide, radon, and fungi were only of inconsequential concentrations in nonresidential buildings and had no determinantal impact on occupant health. On the other hand, carbon dioxide was often found to exceed the health threshold of 1000 ppm in naturally ventilated buildings. In buildings with mechanical ventilation systems, volatile organic compounds, and formaldehyde were found to reach critical levels if the HVAC system is switched off. $PM_{2.5}$ concentration is found to be highly dependent on outdoor concentration, and critical pollution levels were often noted in locations with high outdoor pollution.
Studies often made additional measurements to identify determinants and correlations alongside indoor air pollutants. The analysis of correlations identified building characteristics, equipment, indoor climate, occupant and cleaning activity, building location, proximity of traffic, and country as main influences on indoor air pollution concentration and the outdoor environment.

The second part reviewed the measurement methods for indoor air pollutants in the literature. In general, a differentiation between off-site laboratory and on-site measurements can be made. Laboratory measurement methods are of high cost, complex and stationary and are applicable for legal monitoring. However, laboratory measurement methods can only assess the pollutant load at specified points in time since measurement samples have to be transferred to the laboratory for analysis. On-site measurements, on the other hand, are widely applicable and capable of continuous monitoring on-site; the accuracy is typically lower than laboratory methods and is unsuitable for differentiation between different VOCs. However, on-site measurements can enable live identification of pollution events and ventilation control. On-site measurement methods can be further differentiated by their measurement principle. Three main principles exist : electrochemical, electrical, and optical. In the literature, a detailed assessment of the various measurement principles was made, and specific principles are suitable for different pollutants. However, all of them suffer from longevity, drift and calibration issues. All in all, frequent maintenance and recalibration are key to upholding measurement accuracy. The third section of the literature review explored virtual sensing of indoor air pollutants as an alternative to physical sensing. The literature review indicates that virtual sensing is a viable alternative to physical sensor deployment. Current applications of virtual sensing are HVAC operation monitoring, infiltration rate and zone temperature distribution estimation, zone occupancy, and indoor pollutant monitoring. The literature review showed that virtual sensors could be classified according to their application, modeling method, and data characteristics. In the case of virtual sensing for indoor air pollutant concentration, a black box approach using machine learning methods was found to be most promising. Studies were reviewed that performed virtual sensing using machine learning methods for indoor air pollutants. The reviewed studies compared various machine learning algorithms. Most achieved best results using a multi-layer perceptron model (MLP). Virtual sensors for time series data were also successfully generated in other domains using a long short-term memory (LSTM) model.
The final section of the literature review looked at indoor air pollutant-based demand-controlled ventilation. The literature review showed that demand-controlled ventilation is the only active measure to mitigate indoor air pollutants in the building stock by diluting and filtering pollutants. The primary goals of indoor air pollutant -based DCVs are indoor air quality improvement and energy consumption reduction. DCV is often based on pollutant or occupancy proxies in the built environment since widespread measurements are seldom available. The review identified the main proxies as: humidity, carbon dioxide, facial recognition, and motion detection. Furthermore, the robustness of DCV was assessed by looking at the influence of sensor placement and sensor failures on the ventilation control. The review showed that properly placing the sensors and identifying sensor failures is critical to avoid oversupply or undersupply of fresh air in the zone. Studies looking at the improvements of DCV foremost identified a significant reduction of energy consumed by the HVAC system and the potential to improve air quality while upholding thermal comfort.

3.5.2 Identified Gaps

The literature review allowed to identify several gaps in research regarding the state, measurement, and control of indoor air pollutants in nonresidential buildings.
Many studies examining indoor air pollutant concentrations in nonresidential buildings used too short measurement periods and based their findings and conclusions on a small dataset, thus introducing significant bias. None of the reviewed studies conducted multi-year measurements of indoor air pollutants. Thus, data regarding seasonality or long-time trends is missing in the literature. Furthermore, most reviewed studies performed measurements at an inadequate sampling rate and therefore didn't account for transient responses of indoor air pollutant concentration. Also, despite identifying high spatial resolution, only one study examined the spatial distribution of pollutants in a workshop. Thus, information on the spatial distribution of indoor air pollutants in other nonresidential typologies is missing in the literature.
The review of the physical measurement methods showed that methods applicable for continuous on-site measurement deliver good results compared to reference measurement methods. However, studies showed on-site sensors often suffer from accuracy and reliability issues after prolonged deployment and require frequent maintenance and recalibration of sensors, to avoid a gradual decline in overall measurement performance.
In the third section, virtual sensing is examined as a potential substitute for the physical detection of indoor air pollutants. Past research has focused on using virtual sensing for estimating indoor air pollution levels. Yet, these investigations predominantly concentrate on a single pollutant, whereas a comprehensive assessment of indoor air quality necessitates the evaluation of various pollutants. Moreover, existing studies have developed virtual sensors using data exclusively from a single area within a particular typology, neglecting the potential applicability of these models to different areas and/or typologies. Furthermore, the reviewed studies relied on less than 12 months of measurements to construct their models, leading to considerable bias in these models. Furthermore, no existing virtual sensing methods for indoor air pollutants have incorporated the long short term memory recurrent neural networks, despite their proven effectiveness in other fields for analyzing time-series data.
The review of studies on demand-controlled ventilation based on indoor air pollutant measurements showed that current implementations in the building stock use proxies for occupancy and indoor air pollutants instead of widespread measurements. Furthermore, it is noted that current implementations using sensors are unrobust and are prone to deliver wrong control signals when placed incorrectly or if sensor failures occur.
However, the literature review indicated the high potential of demand controlled ventilation to reduce energy consumption in buildings, while upholding indoor air quality. Therefore, exploring the applicability of virtual indoor air pollutant sensors, that overcome reliability issues of physical sensors, is identified as gap in the literature.

3.5.3 Contribution of this Dissertation

This dissertation contributes to the state of research by tackling the identified gaps in the literature.
In this study, an indoor air pollutant dataset will be created with long-term, high spatiotemporal resolution data from multiple rooms and typologies, thus surpassing current indoor air pollutant datasets used in literature. This dataset will be used to analyze the current status of indoor air pollution in the examined rooms and buildings and serve as training data for virtual indoor air pollutant sensors based on machine learning techniques, with a focus on Long Short-Term Memory networks (LSTMs). The virtual indoor air pollutant sensor should deliver more robust data than physical sensor implementation and enable ubiquitous indoor air pollutant monitoring since the cost and time expenses for deploying and maintaining sensor networks are unnecessary. Finally, the effects of implementing these virtual indoor air pollutant sensors in demand-controlled ventilation will be tested in two case studies and evaluated for potential energy consumption reduction.

Part II

Methods

4 Building a dataset

Certain portions of the textual content, data, figures, and results included in this chapter have been previously published in Gabriel and Auer [2022] and Gabriel and Auer [2023].

The following chapter covers the methods used to build the dataset that is later used in training the virtual sensing models and the case studies as well as analyzing the state of indoor air quality in the examined rooms. This section includes the measurement setup, the measurement locations, the measurement equipment, as well as the procedure to ensure the quality of the dataset. Measurements were performed in multiple rooms and buildings from June 2021 to April 2023. In the following, it is differentiated between the training room, where measurements were used for training and validating the model, and testing rooms, where the model's transferability is tested and whose data is not used for model training.

4.1 Parameters

The following section lists the measurement parameters differentiated by indoor air pollutant measurement parameters , BMS data measurement parameters and outdoor measurement parameters.

4.1.1 Indoor Air Pollutant parameters

Table 4.1 shows the indoor air pollutant parameters to be collected and their respective units and measurement interval. The selection was based on the section 3.1.1 of the literature review, identifying VOC, CO_2 and $PM_{2.5}$ as most relevant indoor air pollutants in non-residential buildings.

Table 4.1: Indoor air pollutant datapoints

Indoor Measurement	**Unit**	**Interval**
VOC	IAQI	10 seconds
CO_2	ppm	10 seconds
$PM_{2.5}$	µg/m^3	10 seconds

4.1.2 Building management system parameters

In addition to measuring indoor air pollutant concentrations a set of BMS data points is defined that are included in building and training the virtual sensors based on machine learning methods. Table 4.2 shows the BMS data parameters to be collected and their respective units and measurement interval.

Table 4.2: BMS datapoints

Indoor Measurement	**Unit**	**Interval**
Temperature	°C	1 minute
Humidity	% RH	1 minute
Air pressure	hPa	1 minute
Illumination	lux	1 minute
Noise-level	dB(A)	1 minute
Window opening state	binary	at Interaction
Equipment power consumption	W	1 minute

4.1.3 Outdoor parameters

Furthermore, outdoor environmental parameters will be collected to be included in the virtual sensing model since the literature review identified a significant indoor-outdoor correlation especially in the case of $PM_{2.5}$. Table 4.3 shows the outdoor metereological and pollution parameters to be collected and their respective units and measurement interval.

Table 4.3: Outdoor Environmental Parameters

Outdoor Measurement	**Unit**	**Interval**
Air temperature	°C	1 minute
Ground temperature	°C	1 minute
Dew point temperature	°C	1 minute
Global and diffuse radiation	W/m^2	1 minute
Humidity	% RH	1 minute
Illumination	lux	1 minute
Air pressure	hPa	1 minute
Precipitation	mm	1 minute
Sunlight hours	hours	1 minute
Wind direction and speed	°, m/s	1 minute
Outdoor $PM_{2.5}$ concentration	$\mu g/m^3$	1 minute

4.2 Measurement Locations

Measurements were performed in two room typologies with high people density and occupation times: four open office rooms and one classroom. The rooms were located in two buildings in Munich. The measurement locations are described below. For each room, the placement of the indoor air pollutant measurement nodes follows the guidelines for monitoring indoor air pollutants of the United States Environmental Protection Agency (EPA). The requirements for the positioning are as follows:

- Installation of the nodes in the breathing zone (1.10 m height)
- More than 0.5 m away from walls, corners, and windows
- More than 1 m away from local pollutant sources and occupants
- Not in front or below air supply units
- Not exposed to direct sunlight

A detailed overview of the positioning of the individual nodes is provided in the following section, showing the floor plan and corresponding sensor location for each examined room.

4.2.1 Building A

Building A is a high-rise office building in the center of Munich with 23 stories and a 130,000 m^2 floor area and accommodates about 2,500 employees. The building is supplied with heating and cooling through thermally activated ceilings (concrete core activation) supplied by groundwater heat pumps. A central mechanical ventilation system supplies the building with fresh air introduced into the room through induction units and extracted through exhaust outlets in the center of the zones. The ventilation system is not designed to supply heating or cooling energy. The ventilation operates at a constant schedule of 1.6 air changes per hour between 5.15 am and 8 pm. In addition to the mechanical ventilation systems, rooms in the lower stories (Office 1, Office 2, Office 4) also have operable windows. All rooms have radiation-controlled shading systems that the occupants can override. The building is near much-frequented roads and railway tracks.

Office 1 Office 1 is located on the third floor of Building A. It has two external façades, which are orientated towards North -West and South-East. The room provides workplaces for about thirty-five employees and features operable windows. Measurements were taken in Office 1 from June 2021 to April 2023 with three independent indoor air pollutant nodes. Office 1 has a floor area of 395 m^2 and a indoor air volume of 1.200 m^3.

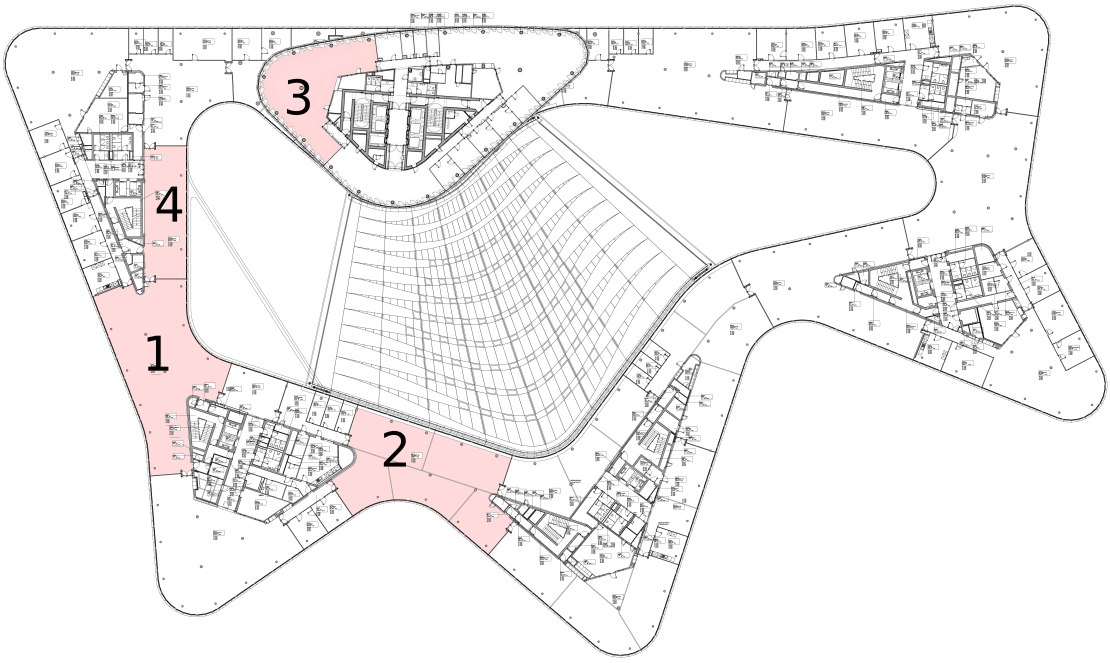
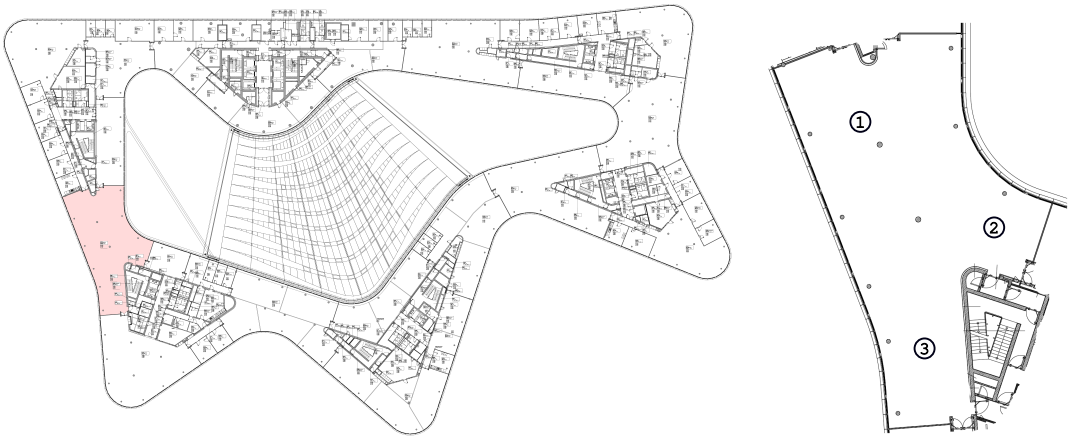


Figure 4.1: Building A floor plan and monitored open office rooms (on multiple floors)



(a) Office 1 location (b) Sensor placement

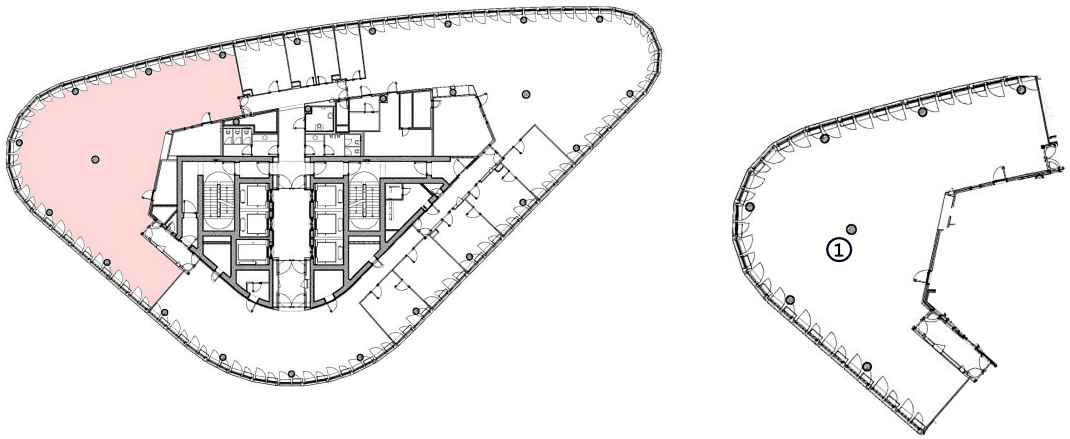
Figure 4.2: Office 1 floor plan and sensor placement

Office 2 Office 2 is located on the fourth floor of Building A. It has two external façades, orientated towards the northeast and south-west, the former looking out on a highly frequented road. The room provides workplaces for about thirty -five employees and features operable windows. Measurements were taken in Office 3 from August 2022 to April 2023. Office 2 has a floor area of 400 m^2 and a indoor air volume of 1.224 m^3.

(a) Office 2 location (b) Sensor placement

Figure 4.3: Office 2 floor plan and sensor placement

Office 3 Office 3 is located on the seventh floor of Building A. It has three external façades, oriented towards the north, north-west, and south-west. The room provides workplaces for about twenty employees and has no operable windows. Measurements have been taken in Office 4 from August 2022 to April 2023. Office 3 has a floor area of 245 m^2 and a indoor air volume of 750 m^3.



(a) Office 3 location (b) Sensor placement

Figure 4.4: Office 3 floor plan and sensor placement

Office 4 Office 4 is located on the third floor of Building A. It has one external façade, which is orientated towards the South-East. The room provides workplaces for ten employees and features operable windows. Measurements were taken in Office 4 from

August 2022 to April 2023 with one indoor air pollutant node. Office 4 has a floor area of 156 m^2 and a indoor air volume of 477 m^3.

(a) Office 4 location (b) Sensor placement

Figure 4.5: Office 4 floor plan and sensor placement

4.2.2 Building B

Building B is a school building in Munich with a 9,000 m^2 total floor area. The building has underfloor heating systems, with heat supplied by a ground-water heat pump for base load and a fossil-based peak load system. The building has no centralized mechanical ventilation. All classrooms have operable windows, with decentralized ventilation systems retrofitted in some.

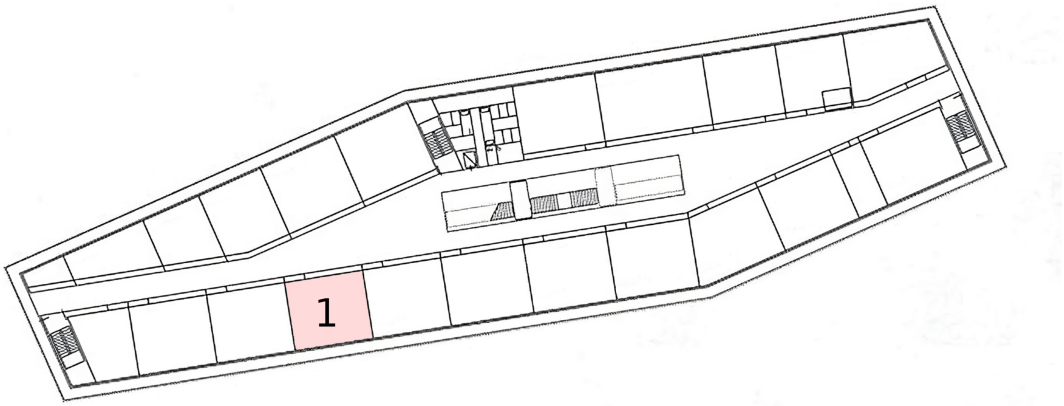
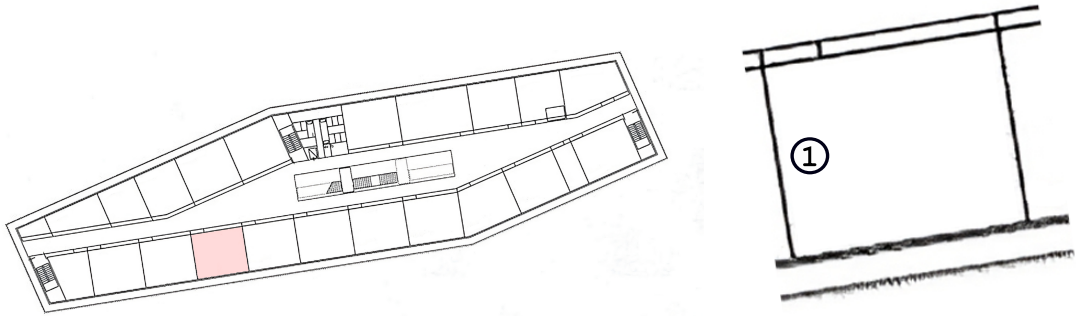


Figure 4.6: Building B floor plan and monitored classroom

Classroom The Classroom in Building B on the second floor is equipped with workplaces for 30 pupils. It has one external façade oriented towards the south and has four operable

windows. Additionally, a decentralized mechanical ventilation system supplies fresh air to the room. The ventilation rate depends on the occupation, and the air is pre-treated by a heat exchanger and a heating register. Measurements were taken in the Classroom from March 2022 to May 2023 with one IAP node deployed.



(a) Classroom location (b) Sensor placement

Figure 4.7: Classroom measurements

4.3 Measurement Equipment

"Measurement equipment was developed for indoor air pollutant (IAP) measurements. The measurement infrastructure for the meteorological and outdoor pollution data was already in place." [Gabriel and Auer, 2023]

4.3.1 Indoor air pollutant measurements

"The IAP nodes are required to measure CO_2 concentrations in parts per million (ppm), [$PM_{2.5}$] concentrations in micrograms per cubic meter ($\mu g/m^3$), and total volatile organic compound concentrations as Indoor Air Quality Index (IAQI). Furthermore, the IAP nodes must achieve continuous, automated measurements with a high sampling rate (10 s) over a prolonged period of time and should account for measurement drift by frequent recalibration. While the initial data sampling rate is 10 s, these measurements will be resampled to a one-minute interval later. This higher sampling rate allows for smoother and more reliable data, as it enables using more data points for each resampled data point. Due to the high volume of data collected, data must be stored centrally rather than locally on the measurement nodes. Therefore, a communication infrastructure supporting high data rates and low latencies was required. [...] No currently available commercial system fulfilled these requirements. Therefore, custom indoor air pollutant nodes (see Figure 4.8) were developed in order to meet the requirements. Sensors were selected based on their evaluation in the literature.
For [$PM_{2.5}$] measurements, the Sensirion SPS30 sensor [...] [was] selected based on its evaluation in previous studies [Demanega et al., 2021]. [Demanega et al., 2021] ascertained a very strong correlation with the reference instrument for [$PM_{2.5}$]. The Sensirion SPS30 utilizes the optical particles counter measurement principle, which has been shown to have good accuracy in measuring [$PM_{2.5}$] of varying diameters [Frederickson et al., 2021]. For VOC measurements, the Sensirion SGP30 [...] sensor [was] chosen. The sensor employs a metal oxide sensing (MOS) element, which is able to detect a wide range of volatile organic compounds through changes in the material's resistance due to chemical reactions with the pollutants. However, due to their broad sensitivity, it is not possible to identify the pollutant concentrations of individual VOCs, which means the output value of these sensors is qualitative. However, [Demanega et al., 2021] evaluated a range of VOC MOS sensors under different pollution events and, in the case of the Sensirion SGP30, performed well compared to reference instruments, thus making it viable for a qualitative evaluation of VOC pollution. For CO_2 measurements, [...] the Sensirion SCD30 [...] [was chosen] due to its proven accuracy [Frederickson et al., 2021]. This sensor uses the optical NDIR measurement principle, which is the common standard in accurately measuring CO_2 concentration [Marinov et al., 2018]. The sensors were connected to a microcontroller, which performs continuous measurements in the defined interval, automatic recalibration, and upload the data to a central database via WiFi connectivity." [Gabriel and Auer, 2023] Table 4.4 gives an overview of the used measurement equipment.

Figure 4.8: Custom-built IAP sensor node

Table 4.4: Overview of the measurement equipment and sensors used.

Property	Sensirion SPS30	Sensirion SGP30	Sensirion SCD30
Parameter	Particulate Matter	Volatile Organic Compounds (VOC)	Carbon Dioxide (CO_2)
Measurement Principle	Optical Particle Counter	Metal Oxide Sensing (MOS)	Optical NDIR
Evaluation Source	Frederickson et al. [2021], Demanega et al. [2021]	Demanega et al. [2021]	Frederickson et al. [2021], Marinov et al. [2018]
Measurement Interval	10 s	10 s	10 s
Use in Literature	Hassani et al. [2023] Kuula et al. [2020]	Alonso et al. [2022] Arsiwala et al. [2023]	Trilles et al. [2021] Toschke et al. [2022]

4.3.2 BMS data measurements

Even though BMS data was available in both buildings, a secondary measurement infrastructure was built for gathering BMS due to data access constraints. Exporting data in the existing BMS systems required a manual export which was adverse to the real-time and online monitoring and would have impeded short development cycles. Therefore, it was decieded to build a secondary BMS system based on the LoRaWAN standard. LoRaWAN is a wireless BUS standard that achieves long-range communication with low power consumption, thus enabling battery-powered nodes. Due to the minimal installation effort and battery-powered nodes, it is applicable as a retrofit solution. The measurment equipment was selected to align with the existing BMS system. Therefore, readily available commercial sensor nodes were chosen. The BMS system recorded measurements at a 1 min interval and allowed for live access and export of the data.

4.3.3 Outdoor meteorological data gathering

Meteorological measurements are collected from openly available interfaces from the meteorological institute of the Ludwig-Maximilian-Universität Munich (LMU). The measurements are taken at Theresienstraße 37 in Munich, less than 4 kilometers from the Office Building in this study and less than 12 kilometers from the school building. Table 4.5 lists the metereological measurements. The measurements are updated at a one-minute interval.

Table 4.5: Measurements from the LMU meteorological station

Measurement	**Height/Depth**
Air Temperature	2.0 m, 30.0 m
Wet Bulb Temperature	2.0 m, 30.0 m
Dew Point	2.0 m, 30.0 m
Relative Humidity	2.0 m, 30.0 m
Windspeed	30.0 m
Soil Temperature	50 cm, 20 cm, 10 cm, 5 cm, 2 cm
Global Radiation	Surface
Diffuse Radiation	Surface
Downward Terrestrial Radiation	Surface
Sunshine Duration	Surface
UV-Index	Surface
Air Pressure (515 m altitude)	Surface
Air Pressure (Sea level)	Surface
Wind Direction (30 m height)	Surface
Precipitation Current	Surface
Precipitation Accumulated	Surface
Precipitation Type	Surface

4.3.4 Outdoor air pollutant data gathering

Data from the citizen science project *luftdaten.info* is used for outdoor $PM_{2.5}$ measurements. *Luftdaten.info* is an initiative built on the voluntary participation of individuals who build and operate $PM_{2.5}$ sensors in the outdoor environment to gain high spatial resolution outdoor pollution measurements. In, Munich, more than 130 measurement points are currently in operation. Each sensor reports at a one-minute measurement interval.
The initiative is backed by several scientific and administrative institutions, which evaluate the sensors' accuracy and help select appropriate technology (*luftdaten.info/evaluation/*). Table 4.6 shows the measured data points of each unit. This study uses data from luftdaten.info to integrate outdoor $PM_{2.5}$ concentration values into the model. The

Table 4.6: Outdoor Pollution Measurements from Citizen Science Project

Measurement	**Unit**
Particulate Matter (2.5 μm)	$\mu g/m^3$
Particulate Matter (10 μm)	$\mu g/m^3$
Temperature	°C
Humidity	% RH
Air pressure	hPa
(Optional) Noise	dB

measurement location is less than 300 meters in distance in the case of the office building as well as less than 500 meters for the school building.

4.4 Quality control

Since multiple indoor air pollutant nodes would be deployed, it was important to reduce sensor bias. Therefore, a cross-calibration scheme was introduced in this study. Cross-calibration is a method used to reduce sensor bias and improve accuracy by comparing the readings of individual sensors to a chosen reference sensor. In this case, one sensor was selected as the reference, and all other sensors were calibrated to perform accordingly. This approach ensures consistency among the sensor readings. The cross-calibration procedure was conducted over a 24 h period, during which a wide range of environmental conditions were introduced to test sensor response over the entire measurement range. "Based on the [...] [gathered data, calibration curves for each individual sensor are generated using regression analysis outputting] a polynomial function for each sensor and measurement. The polynomial function was used to calibrate the raw measurement values. Due to the SPS30s underreporting of [$PM_{2.5}$] concentration, a calibration function was derived from the measurement [data] in Demanega et al. [2021] comparing SPS30 and [a reference instrument]. A regression analysis determined a linear function between SPS30 and [reference] measurements. The derived linear function was then applied to all SPS30 raw measurements in this study" [Gabriel and Auer, 2022] to remedy the underreporting.
Figure 4.9 show an exemplary calibration function for VOC measurements derived from measurement data.

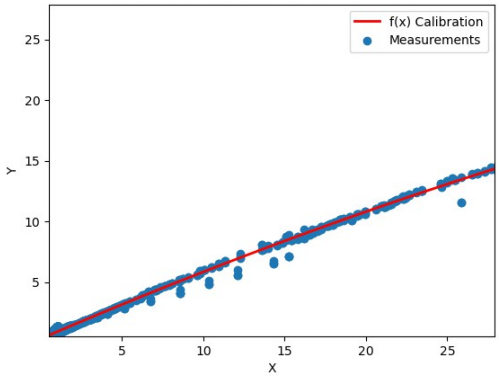


Figure 4.9: Exemplary calibration function derived from measurement data

5 Virtual sensing of indoor air pollutants

Certain portions of the textual content, data, figures, and results included in this chapter have been previously published in Gabriel and Auer [2022] and Gabriel and Auer [2023].

This study utilizes machine learning (ML) models that have been trained on collected measurement data to predict indoor air pollutant concentrations. The models take inputs from the Building Management System (BMS), outdoor meteorological data, and outdoor pollution data. The following sections provide a detailed explanation of the methods used. Section 5.1 outlines the steps taken to preprocess the data in preparation for training the machine learning models. Section 5.2 describes the process of training the models, while section 5.3 focuses on evaluating the performance of the models. Section 5.4 provides an overview of the transferability tests conducted.
For the purposes of this study, the data from Office 1 will be used to train the model, fine-tune its hyperparameters, and evaluate and select the best model. In the subsequent section, Office 1 will also be referred to as the "training room". Office 2 to 4 and the Classroom will not be used for training the model, but rather serve as unseen test environments for the virtual sensing model. These rooms will be referred to as the "testing rooms". The study will take two approaches. The first approach involves training and evaluating the model in the training room, and then testing its transferability to another room without making any adjustments to the original model. The second approach involves training and evaluating the model in the training room, and then utilizing transfer learning to adapt the model to different environmental conditions in other rooms, buildings, and typologies. A graphical representation of the process is provided in figure 5.1.

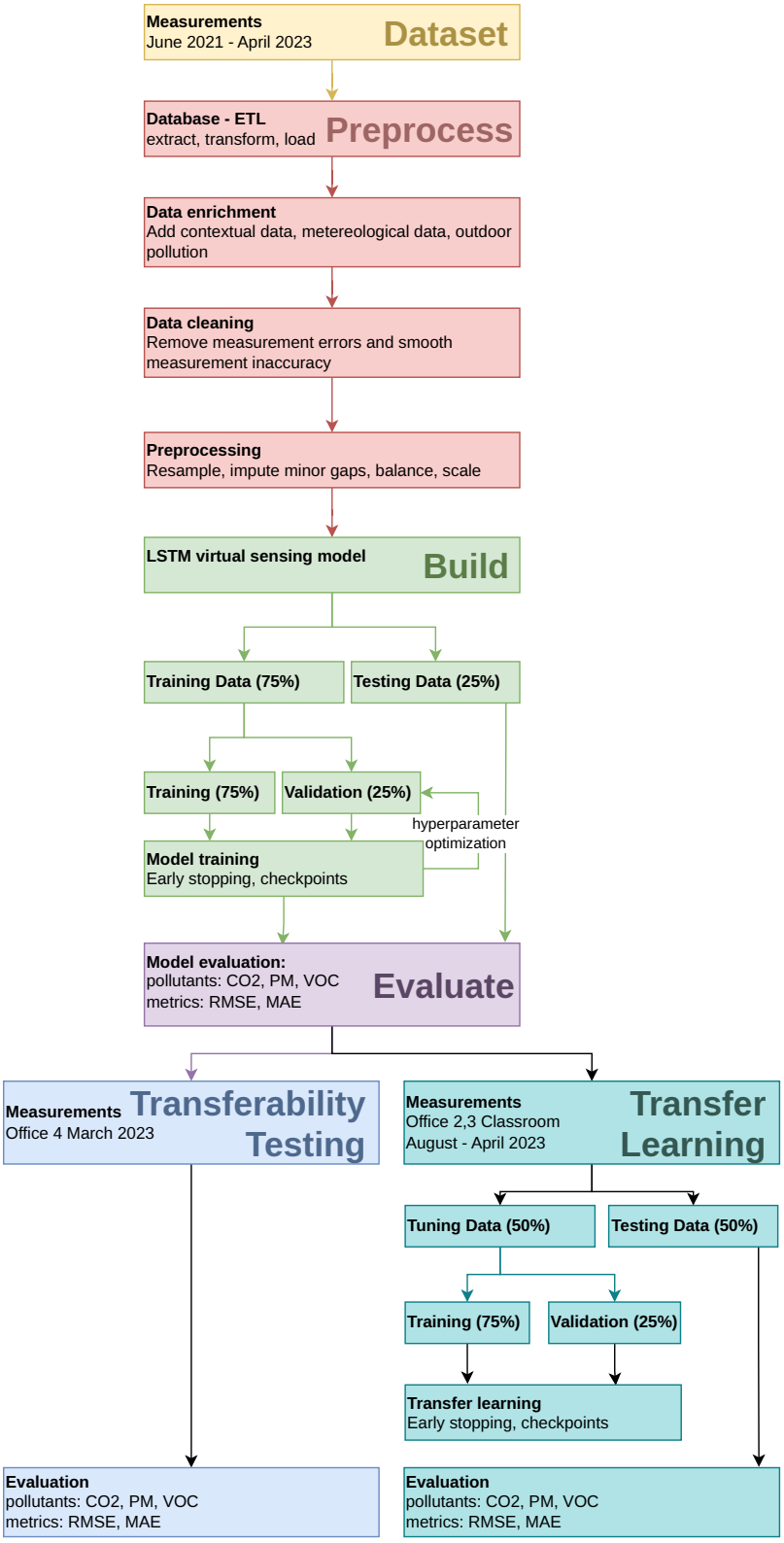


Figure 5.1: Graphical illustration of the methods. Right: approach a) - model training and transfer learning; Left: approach b) model training and transferability testing

5.1 Preprocessing

The dataset's creation is described in the previous method chapter of this dissertation. The data from Office 1 and outdoor meteorological and pollution data will be used for model training and evaluation. Measurement data from Office rooms 2 to 4 and the Classroom will be used in model evaluation, transferability testing (unseen transfer), and a transfer learning approach.
The following section outlines the steps taken to preprocess the raw datasets for machine-learning model input. These steps include filtering relevant data, handling missing or corrupted values, normalizing the data, and splitting the data into training, validation, and test sets. Figure 5.1 provides a graphical summary of the data preparation process.
"The initial data preparation involved extracting measurement data from IAP and the BMS node from the database and transforming the data from a long to a wide format. [The transformed] data was then loaded into a Pandas data frame for further processing." [Gabriel and Auer, 2023] Pandas, a widely used Python library for data analysis, was utilized for its data storage and manipulation capabilities in preparation for machine learning tasks.
"The available measurement data was enriched by adding contextual and outdoor environmental data." [Gabriel and Auer, 2023] Contextual data is date and time information, with hours and days encoded as continuous sinusoidal values. Furthermore, binary tags were added to indicate workdays, weekends, holidays, and seasons. Additionally, information on the HVAC operation schedule, room size, and number of occupants were integrated into the model.
Furthermore, outdoor environmental data from a local meteorological station was included in the model. This data encompassed air temperature, ground temperature, dew point temperature, global and diffuse radiation, humidity, illumination, air pressure, precipitation, sunlight hours, wind direction and speed, and outdoor $PM_{2.5}$ concentration.
It was observed that measurements taken immediately after power cycling the nodes, such as after a power outage, exhibited temporarily elevated temperature and humidity values. To prevent model bias, measurements up to 15 minutes after a power cycle were excluded from the dataset. Additionally, random measurement fluctuations caused by sensor inaccuracies were programmatically smoothed out from the dataset using an exponential moving window function.
The final preprocessing steps involved optimizing the dataset for machine learning. To ensure continuous and evenly spaced measurement intervals, the measurement data was resampled to a one-minute frequency. Small intervals of missing data (up to 15 minutes) were imputed with the mean value of each measurement. Finally, the datasets were balanced and normalized using a min-max scaler for each feature.

5.2 Machine Learning Algorithms

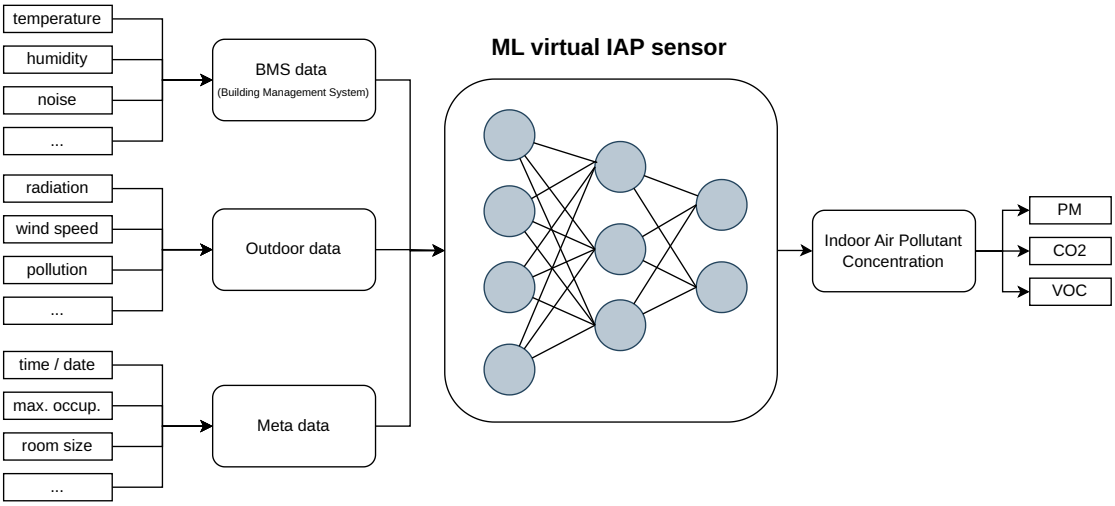


Figure 5.2: Visual representation of the study definition

Various machine learning (ML) methods were tested in the development of virtual indoor air pollutant sensors for this dissertation. Three approaches were compared: long-short term memory recurrent neural networks (LSTM), multi-layer perceptron neural networks (MLP), and stochastic gradient descent (SGD). The selection of these algorithms is based on the literature review in the state of research chapter, where they were identified as the most commonly used algorithms in similar studies. In the following section, the configuration of these algorithms is explained.

The LSTM was implemented with two hidden layers. The input data is fed into the LSTM as a three-dimensional tensor. "The first dimension represent[s] the length of the input variables (e.g., temperature, humidity [...]), the second dimension [represents] the lookback period (i.e., the number of past timesteps), and the third dimension represents the batch size, which indicates the number of input sequences processed concurrently during training and inference." [Gabriel and Auer, 2023] The hyperparameters of the LSTM, including the learning rate, batch size, lookback period, and number of neurons in the hidden layers, were determined through hyperparameter optimization.

The setup of the SGD and MLP models was slightly different because their model architectures do not consider past timesteps. Instead of providing a three-dimensional input tensor, the data was provided as a single-dimensional input vector with the length equal to the number of input variables. Since batch size and lookback period are not relevant for SGD and MLP models, these hyperparameters were not optimized for these models.

The hyperparameters of the models were optimized using a Bayesian optimization algorithm to minimize the model error. "[Furthermore,] an early stopping function was implemented to prevent model overfitting by monitoring the validation loss and terminating model training if the validation loss did not improve for five consecutive

[steps]."[Gabriel and Auer, 2023] The overall training time for the LSTM model was 28 minutes on a GPU. The MLP and SGD models were trained on the CPU and required 23 and 3 minutes, respectively, to complete the training procedure. An overview of the model input and output is provided in Table 5.1.
The machine learning model was trained using data collected from Office 1. A cross-validation scheme was employed to ensure that the models could generalize and predict indoor air pollutant concentrations. This involved reserving 25% of the data for testing purposes and using the remaining data to train the models. The training data was further divided into a training set (75%) and a validation set (25%). The validation set was used to trigger the early stopping algorithm, prevent overfitting, determine the best epoch, and perform hyperparameter optimization. Both the LSTM and MLP models were configured with two hidden layers. This configuration, based on the literature review of similar studies in the state of research chapter, provided a sufficient model depth to capture the complex relationships between the input and output variables while maintaining performance and avoiding overfitting. The model architecture and configuration of the LSTM model are shown in Figure 5.3.
Figure 5.4 displays the training loss and validation loss of the LSTM model. These

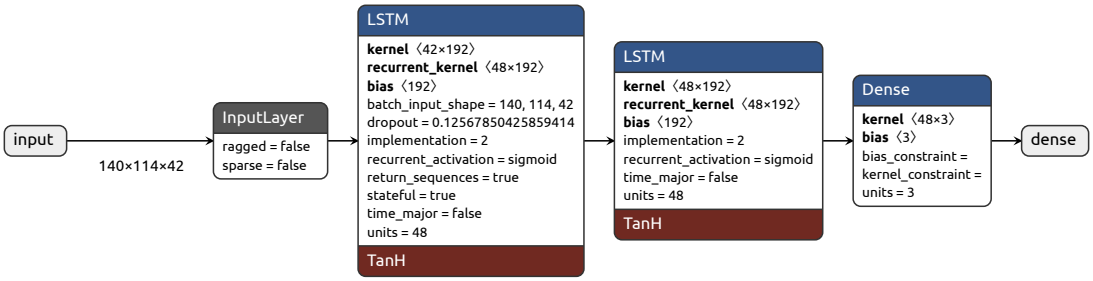


Figure 5.3: LSTM model architecture (own representation, produced using Netron app)

loss functions serve as metrics to evaluate the performance of machine learning models. The training loss measures how well the model fits the training data by quantifying the similarity between predicted outputs and actual target values. Conversely, the validation loss assesses the generalization capability of the model by evaluating its performance on a separate validation dataset that was not used during training. The objective is to minimize both the training and validation losses, ensuring that the model learns patterns from the data without overfitting. Any discrepancies observed between the training and validation losses may indicate potential overfitting or underfitting issues.
The figure 5.4 shows that the training loss consistently decreases over time. Conversely, the validation loss initially decreases until epoch 8. Consequently, training is halted at epoch 11 due to the implementation of the early stopping function. The input and output parameters of the virtual sensing model are detailed in Tables 5.1.

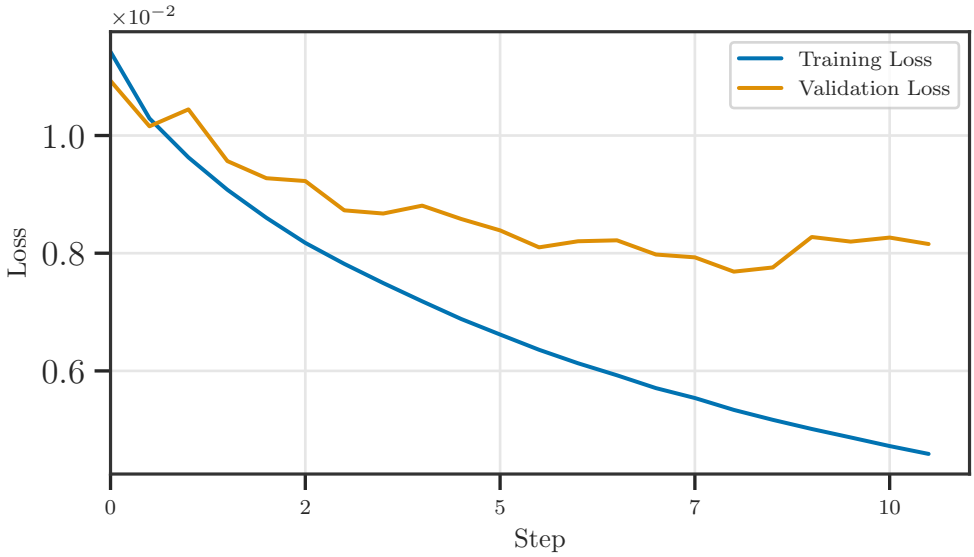


Figure 5.4: Training and validation loss of the LSTM model

Table 5.1: Input and output parameters of the virtual sensing model

Feature	Description	Group
month_sin	Continuous sinusoidal encoding of month	Meta
hr_sin	Continuous sinusoidal encoding of hour	Meta
day_sin	Continuous sinusoidal encoding of day	Meta
workday	Boolean tag for workdays	Meta
weekend	Boolean tag for weekends	Meta
holiday	Boolean tag for holidays	Meta
season	Boolean tags for each season	Meta
room_size	Size of the room	Meta
occupants	Occupant density	Meta
temp	Outdoor air temperature	Outdoor
ground_temp	Outdoor ground temperature	Outdoor
dew_point_temp	Outdoor dew point temperature	Outdoor
global_rad	Outdoor global radiation	Outdoor
diffuse_rad	Outdoor diffuse radiation	Outdoor
humidity	Outdoor humidity	Outdoor
illumination	Outdoor illumination	Outdoor
air_pressure	Outdoor air pressure	Outdoor
precipitation	Outdoor precipitation	Outdoor
wind_dir	Outdoor wind direction	Outdoor
wind_speed	Outdoor wind speed	Outdoor
particulate_matter	Outdoor $PM_{2.5}$ concentration	Outdoor
indoor_temp	Indoor air temperature	Indoor
indoor_humidity	Indoor humidity	Indoor
indoor_air_pressure	Indoor air pressure	Indoor
indoor_illum	Indoor illumination	Indoor
noise_level	Indoor noise level	Indoor
window_state	State of window (open/closed)	Indoor
power_consumption	Power consumption of equipment	Indoor
hvac	Boolean tag for HVAC operation	Indoor
pm	$PM_{2.5}$ concentration	**Output**
co2	CO_2 concentration	**Output**
voc	Volatile organic compound concentration	**Output**

5.3 Evaluation

The model predictions were evaluated using a separate testing dataset. The mean absolute error (MAE) and root mean squared error (RMSE) metrics were used to quantify the performance. These metrics were calculated for each model, pollutant, and room individually. The MAE and RMSE were chosen as they are commonly used in model performance evaluation [Willmott and Matsuura, 2005]. The R2 metric, although widely used in literature, was not considered in this study as it is not suitable for nonlinear models such as LSTMs or MLPs with multiple hidden layers, as pointed out by Spiess and Neumeyer [2010] and Sapra [2014]. The MAE and RMSE metrics are not dimensionless and are expressed in the units of the target variable being evaluated. The MAE represents the mean absolute difference between the predicted and true values for all tested time steps. The RMSE, on the other hand, incorporates a quadratic component in its calculation, giving more weight to larger errors compared to smaller ones [Willmott and Matsuura, 2005]. Consequently, the Mean Absolute Error (MAE) indicates the overall error in target units, while the Root Mean Square Error (RMSE) indicates the number of high deviations. Smaller values for both RMSE and MAE signify a better fit. The RMSE is calculated using the formula:

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{N}} \tag{5.1}$$

The MAE is calculated using the formula:

$$MAE(y, \hat{y}) = \frac{\sum_{i=0}^{N-1}|y_i - \hat{y}_i|}{N} \tag{5.2}$$

5.4 Transferability Testing (without model tuning)

"To evaluate the model's ability to predict indoor air pollutant concentrations in other rooms and environments, the trained and assessed models [are] transfered to an unseen office room (Office 4) in the same building. [This room] has a different layout, occupancy patterns, density, and orientation." [Gabriel and Auer, 2023] The same Building Management System (BMS) data points that were used to train the LSTM model in Office 1 are available in Office 4. The outdoor meteorological and pollution data are retrieved from the same source, as both rooms are in the same building. "The measurements of the [Indoor Air Pollution] nodes are solely used for evaluation in Office 4. The trained model is used as is [and was not trained on the data of Office 4]. The model inputs, as specified in Table 5.1, [are] provided by the BMS node as well as outdoor and metadata. The model then predicts the indoor air pollutant concentrations for each time step (1 minute). In the evaluation, the predicted values are then compared to the actual measurements of the IAP nodes for March 2023." [Gabriel and Auer, 2023] As previously, the prediction metrics, MAE and RMSE, are calculated.

5.5 Transfer learning (with model tuning)

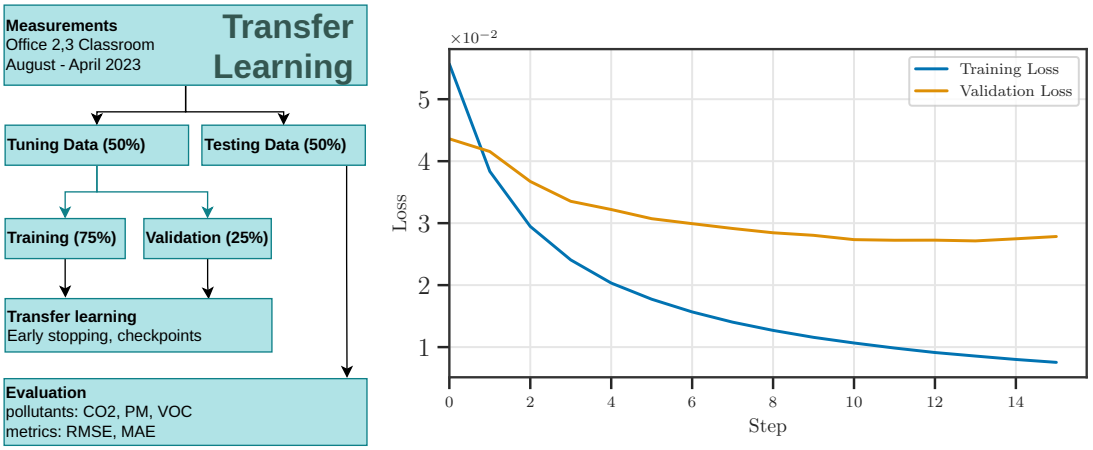


Figure 5.5: Transfer learning approach (left) and training model loss (right)

In addition to testing the transferability of the trained model without adapting the original model to its new environment, a transfer learning approach is also tested. This approach takes the original trained model and adapts it to the new environment using a short-time measurement dataset from the new room for a fine tuning of the model. Transfer learning was tested in three different environments: Office 2, Office 3, and the Classroom. In each environment, the performance of the pre-trained model was evaluated using a dataset of indoor air pollutant concentrations that were measured. Prior to conducting the tests, the pre-trained model underwent a fine-tuning process using measurement data collected over a two-month period from each environment. This fine-tuning process allowed the model to adapt to the unique characteristics of each environment. The specific details of the transfer learning process are described below. To fine-tune the pre-trained model, a cross-validation scheme was employed. Two months of data were set aside for testing and were excluded from the model training process. The initial two months of data were divided into a training set (75%) and a validation set (25%), which were then used to fine-tune the model. To prevent the fine-tuning process from overwriting the weights of the pre-trained model, a lower learning rate was selected for the optimization algorithm. Additionally, the weights of the first LSTM and MLP layer were kept fixed, as suggested by previous studies on transfer learning. Furthermore, an early stopping function was implemented to halt the training process if the model's performance on the validation set failed to improve after five iterations. As mentioned before, the predictions were evaluated against measured ground truth data using selected evaluation metrics for the testing period of the transfer learning. The training and validation loss of the fine-tuning process for the LSTM model is illustrated in the figure. The training loss consistently decreased over the epochs, while the validation loss initially decreased but showed no further improvement after epoch 10. Consequently, the early stopping function was triggered after epoch 15.

Part III

Results

6 Dataset characteristics

In the following chapter, the indoor air pollutant dataset will be analyzed for each examined room. Each section will look into one room, and a final comparison section will compare all rooms.

6.1 Office 1

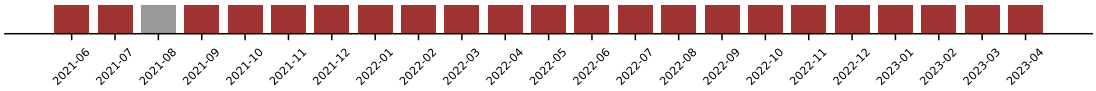


Figure 6.1: Data collection period Office 1

Data was collected in Office 1 from June 2021 to April 2023. Due to changes to the data collection system, no data was collected during August 2021. However, occupancy was low during this period due to the summer holidays. Data was gathered from multiple measurement nodes with a measurement interval of 10 seconds. All in all, resulting in 5.3 million measured time steps for 97 individual data points spanning indoor and outdoor measurements and metadata. All in all, 514 million data points were collected in Office 1.

6.1.1 Statistical analysis

Key statistical values were calculated for the indoor measurements, including indoor air pollutant measurements (CO_2, VOC, $PM_{2.5}$) and BMS data (illumination, pressure, sound, temperature, humidity). For each measurement, mean, standard deviation (std), minimum (min), 25 percentil (25%), 50 percentil (50%), 75 percentil (75%) and maximum (max) were calculated. The table summarizes the statistical values for indoor measurements in Office 1. Several assessments can be made based on the data. **CO_2:** The data shows a mean of CO_2 only 83 ppm over outdoor concentration and a small standard deviation of 36.8 ppm. Furthermore, percentile data shows that up to 75. percentile, all data lies below 500 ppm, and only outliers reach higher concentrations. This can be attributed to the high continuous air change rate of 1.6 air changes per hour and the continuously low occupancy due to the increase in home office due to the Covid pandemic.
$PM_{2.5}$: Similar to CO_2, $PM_{2.5}$ concentration is very low for at least 75% of the time, with a very low mean of 0.8 $\mu g/m^3$. However, peaks up to 70 $\mu g/m^3$ show that seldom high pollution events arise.
VOC: VOC are evaluated using the IAQI (indoor air quality index) with a typical range

Table 6.1: Statistical analysis for indoor measurements Office 1

	CO_2	VOC	$PM_{2.5}$	illu.	press.	sound	temp.	hum.
mean	483.0	51.9	0.8	3.3	955.6	49.9	22.1	36.4
std	36.8	39.0	1.0	5.1	13.6	6.2	0.7	11.4
min	400.0	0.0	0.0	0.0	758.6	32.1	15.2	4.6
25%	457.8	19.5	0.2	0.0	951.3	46.4	21.6	27.7
50%	474.9	47.0	0.4	1.4	956.2	48.3	22.1	34.0
75%	499.2	77.0	1.0	5.1	960.7	51.6	22.6	44.3
max	1468.6	249.5	69.8	222.3	1203.7	96.0	26.7	94.1

from 0 (clean air) to 350 (extremely polluted air). VOC shows higher variability than the other two air pollutants in Office 1; however, no critical thresholds have been reached up to the 75 percentile. However, peaks up to 250 IAQI show that seldom high pollution events arise.

Temperature and humidity: Data for temperature and humidity show a very static indoor environment with minimal variations. Percentiles for temperature and humidity are close, and the standard deviation is low in both cases. A possible explanation is the concrete core heating and cooling system that provides stability through high exposed thermal mass. Furthermore, low occupancy and low equipment load lead to small internal loads.

6.1.2 Time-series analysis

The temporal distribution of pollutant events is analyzed in the following section. The

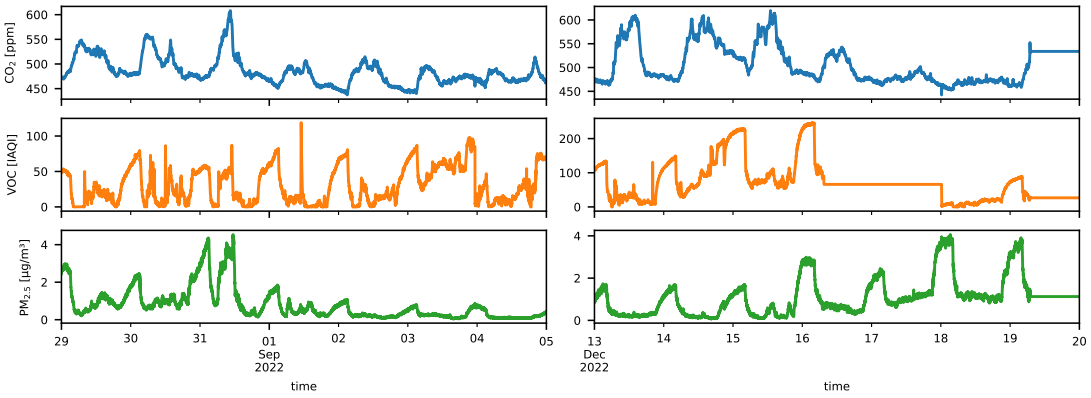


Figure 6.2: Exemplary summer (left) and winter (right) week time series plots for indoor air pollutants in Office 1

figure shows time series plots for each pollutant for an exemplary winter and summer week. It can be seen that CO_2 follows a strong daily frequency for weekdays determined

by occupancy and levels out during weekends. This can be observed both in the winter and summer weeks. In the case of $PM_{2.5}$, a daily pattern is also identifiable, especially for the winter case. Peaks in $PM_{2.5}$ concentration are situated during nighttime hours when the HVAC system is switched off. VOCs show less of a pattern for both winter and summer cases. In the winter, an increase during nighttime hours can be examined, which rapidly decreases when the HVAC is turned on at 5:15. Apart from this, VOC shows sharp pollution peaks for both winter and summer cases during the daytime.

Seasonality In the next plots, the seasonality of the pollutants is examined. Seasonality means a pattern that can be observed over a certain time frame. The observed time frame here is a year. For each month, a boxplot is calculated. The box represents the middle 50% of the data, meaning data from 25. to 75. percentile. The whiskers extend to all data points within 1.5 times the range between 75. and 25. percentile. The line in the middle of the box signifies the median (50. percentile) of the data. The figure

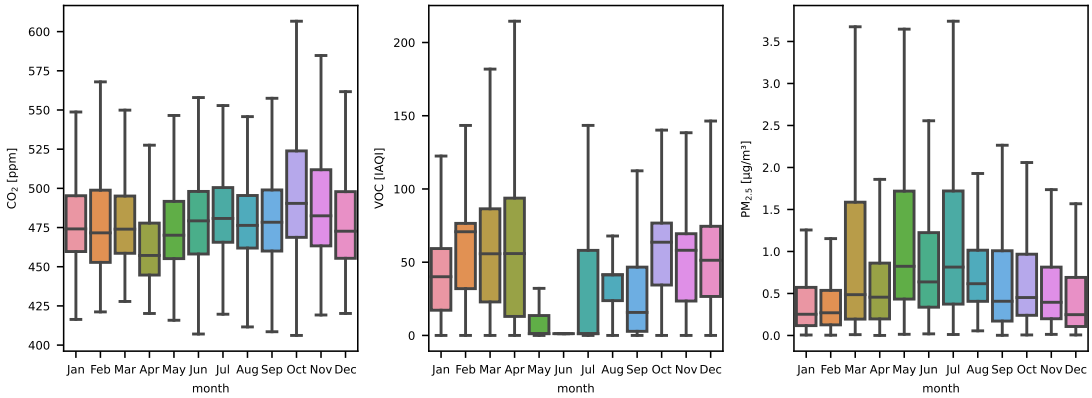


Figure 6.3: Monthly seasonality boxplots for indoor air pollutants in Office 1

shows the seasonality boxplots for the three indoor air pollutants. It can be noted, that CO_2 concentration is relatively constant throughout the months, with slightly higher variability during the winter months. VOCs show higher concentrations and higher volatility towards high concentrations during winter months. VOC concentrations are very low during May-August. This could be attributed to increased natural ventilation at higher outdoor temperatures since the main source of VOC is inside. In the case of $PM_{2.5}$, the case is reversed to that of VOC. In the spring and summer months, volatility and concentrations are higher than in the winter. This can again be attributed to natural ventilation since the main source of $PM_{2.5}$ is outdoors.

6.1.3 Pollution levels

In the following, the pollution levels are compared to WHO limits, and the number of overshoot hours is calculated for each pollutant. The figure shows the number of hours during which a pollutant exceeds the limits set by the WHO. In the case of CO_2, nearly

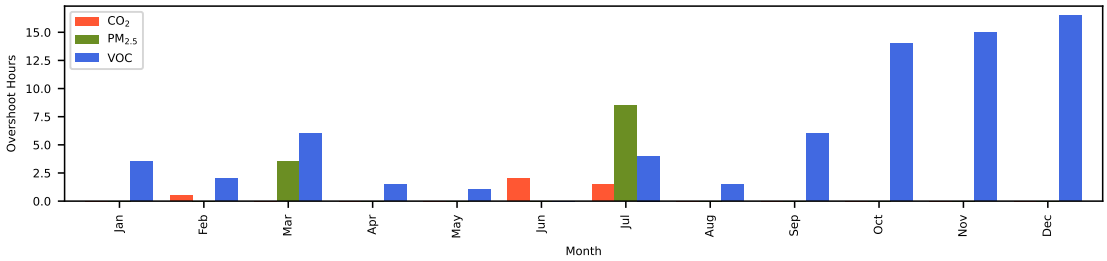


Figure 6.4: Monthly overshoot hours for indoor air pollutants during presence in Office 1

no overruns are recorded. The same can be recorded for most months for $PM_{2.5}$. Only the months of March and July present multiple hours during which the WHO threshold is reached. In the case of VOC, more hours of pollution can be examined, especially during winter.

6.1.4 Correlation analysis

Furthermore, the correlations between the different measurements are analyzed. These include the correlation between indoor pollutant measurements and BMS data, the correlation between indoor pollutant measurements and outdoor measurements, metadata and indoor pollutant measurements, and a time-related analysis. All correlation coefficients used in the following section are calculated using the Pearson correlation coefficient. Correlations >0 .7 or <-0.7 will be interpreted as strong positive or negative correlations. A moderate positive or negative correlation falls within the range of 0.3 to 0.7 (positive) or -0.7 to -0.3 (negative). On the other hand, a correlation is considered weak when it is in the range 0.1 to 0.3 or -0.1 to -0.3. Everything between -0.1 and 0.1 is interpreted as little or no correlation.

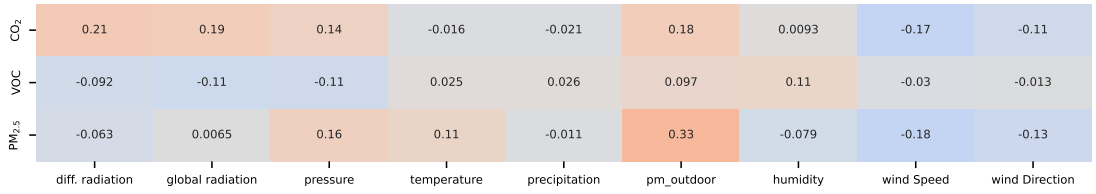


Figure 6.5: Correlation matrix: indoor air pollutants - outdoor measurements in Office 1

Indoor - Outdoor correlation The correlation matrix shows the Pearson correlation between different outdoor measurements and indoor air pollutants . In the case of CO_2, weak positive correlations could be identified for radiation, pressure, and outdoor $PM_{2.5}$ concentration, and weak negative correlations for wind speed and direction.
Regarding VOC, correlations towards outdoor measurements are generally very weak since sources are primarily indoor sources. However, a weak positive correlation was found

for humidity and weak negative correlations for global radiation and pressure. $PM_{2.5}$ has a moderate positive correlation with outdoor $PM_{2.5}$ concentration. Furthermore, it has a weak positive correlation towards pressure and temperature and a negative towards wind speed and direction.

Presence - Time correlation The following section examines the dependence between time, weekday, and pollution concentration. The heatmap shows the hour on the y-axis

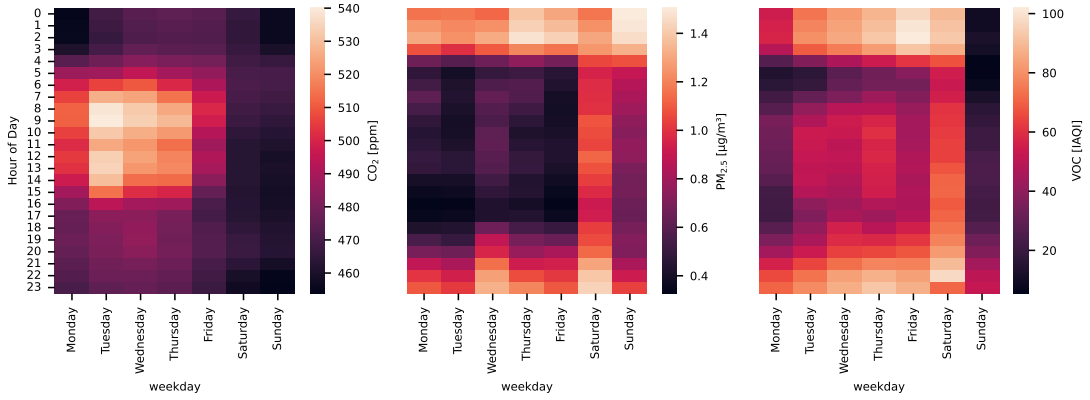


Figure 6.6: Time dependant indoor pollutant concentration in Office 1

and the weekday on the x-axis. The color of the cells is based on the mean pollutant concentration during the specific hour and weekday for the whole measurement period. The left plot visualizes the dependence of CO_2 on time and weekdays. A strong correlation is visible, with maximum CO_2 concentration during weekday morning and early afternoon hours, with low concentration during night and weekends. Furthermore, Friday and, to a reduced degree, Monday show a reduced CO_2 concentration due to the prevailing home office.
The middle plot shows concentration of $PM_{2.5}$ in regards to time and weekday. A reduced $PM_{2.5}$ concentration is recorded between 5 am and 8 pm on weekdays, which overlaps 1:1 with the ventilation schedule. During nights and weekends, $PM_{2.5}$ concentration gradually builds up.
The third plot on the right shows the time and weekday dependence of VOC concentration. Again VOC concentration is strongly influenced by the ventilation schedule and decreases between 5 am and 8 pm during weekdays. However, VOC concentration is elevated, especially during days with higher occupancy (Tuesday, Wednesday, and Thursday). Analog to $PM_{2.5}$, VOC tends to build up during night and weekend periods, where no ventilation is present to disperse the pollutants.

Metadata correlation The following section explores the correlation between collected metadata and indoor air pollutants. The correlation matrix above shows the Pearson correlation coefficients for indoor air pollutants (y-axis) and metadata records (x-axis); the color of the cells corresponds to the correlation. In the case of CO_2, a moderate

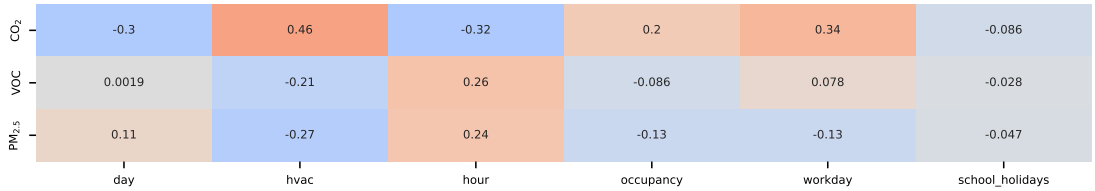


Figure 6.7: Correlation matrix: indoor air pollutants - metadata in Office 1

positive correlation can be determined for the HVAC operation and for the tag indicating whether a day is a workday. Furthermore, a moderate negative correlation was identified for the hour and the day of the week. Furthermore, a weak positive correlation was identified for room occupancy. In the case of VOC, a weak positive correlation is identified for the hour and a weak negative correlation for the HVAC operation. Apart from this, little or no correlation was identified for the day, occupancy, workday, and school holidays for VOC. $PM_{2.5}$ showed a weak positive correlation for the hour and a weak negative correlation to the HVAC operation schedule, room occupancy, and workday indicator.

BMS data correlation The following section explores the correlation between indoor BMS data and indoor air pollutants. The correlation matrix shows the Pearson correlation

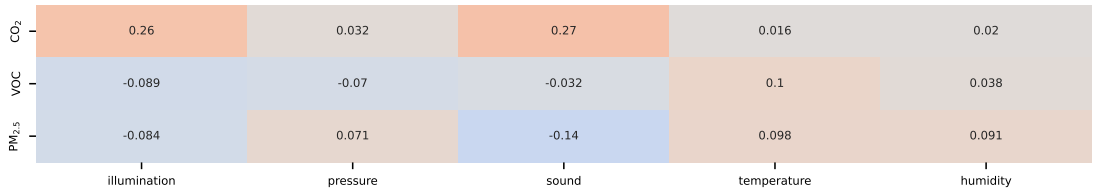


Figure 6.8: Correlation matrix: indoor air pollutants - indoor measurements in Office 1

coefficients for indoor air pollutants (y-axis) and BMS data (x-axis) with cells colored to their corresponding correlation value. CO_2 shows a weak positive correlation towards illumination and sound while showing little correlation towards pressure, temperature, and humidity. VOC shows a weak positive correlation to temperature and little to no correlation to all other indoor BMS data. In the case of $PM_{2.5}$, a weak negative correlation is determined for sound and little correlation to other BMS data.

6.2 Office 2

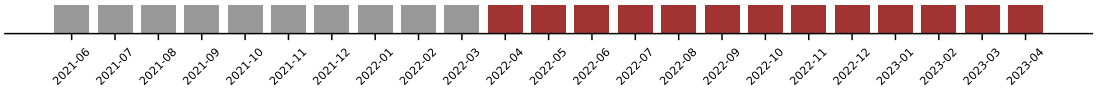


Figure 6.9: Data collection period in Office 2 (red) vs total collection period (grey)

Data was collected in Office 2 from August 2022 to April 2023. Data was gathered from a single measurement node with a measurement interval of 60 seconds. All in all, resulting in 528 thousand measured time steps for 86 individual data points spanning indoor and outdoor measurements and metadata. All in all, 45 million data points were collected in Office 2.

6.2.1 Statistical analysis

Key statistical values were calculated for the indoor measurements, including indoor air pollutant measurements (CO_2, VOC, $PM_{2.5}$) and BMS data (illumination, pressure, sound, temperature, humidity). For each measurement, mean, standard deviation (std), minimum (min), 25 percentil (25%), 50 percentil (50%), 75 percentil (75%) and maximum (max) were calculated.

The table summarizes the statistical values for indoor measurements in Office 2. Several

Table 6.2: Statistical analysis for indoor measurements in Office 2

	CO_2	**VOC**	**$PM_{2.5}$**	**illu.**	**press.**	**sound**	**temp.**	**hum.**
mean	504.8	113.8	4.6	70.2	954.6	64.1	21.7	34.1
std	44.3	67.0	1.9	107.8	8.2	0.8	1.0	10.1
min	400.0	17.0	3.0	0.0	922.0	64.0	20.0	15.0
25%	473.0	54.0	3.0	0.0	950.2	64.0	21.0	25.5
50%	499.0	100.0	4.0	10.0	955.3	64.0	21.3	32.5
75%	529.0	164.0	6.0	114.0	959.6	64.0	22.0	42.0
max	781.0	500.0	17.0	6211.0	973.9	90.0	28.1	62.0

assessments can be made based on the data.

CO_2: The data shows a mean of CO_2 only 104 ppm over outdoor concentration and a small standard deviation of 44.3 ppm. Furthermore, percentile data shows that up to 50. percentile all data lies below 500 ppm, and the maximum concentration only reaches 781 ppm.

$PM_{2.5}$: $PM_{2.5}$ concentration has a mean of 4.6 $\mu g/m^3$, significantly higher than mean concentrations in office 1. However, compared to Office 1, variability is smaller, and peaks only reach up to 17 $\mu g/m^3$, which is still acceptable for short periods.

VOC: VOC shows higher variability than the other two air pollutants; however, up to the 75 percentile, no critical thresholds are reached. However, peaks up to 500 IAQI

show that seldomly extreme pollution events arise.
Temperature and humidity: Data for temperature and humidity show a very static indoor environment with minimal variations. Percentiles 25 - 75 for temperature and humidity are close to each other, and the standard deviation is low in both cases.

6.2.2 Time-series analysis

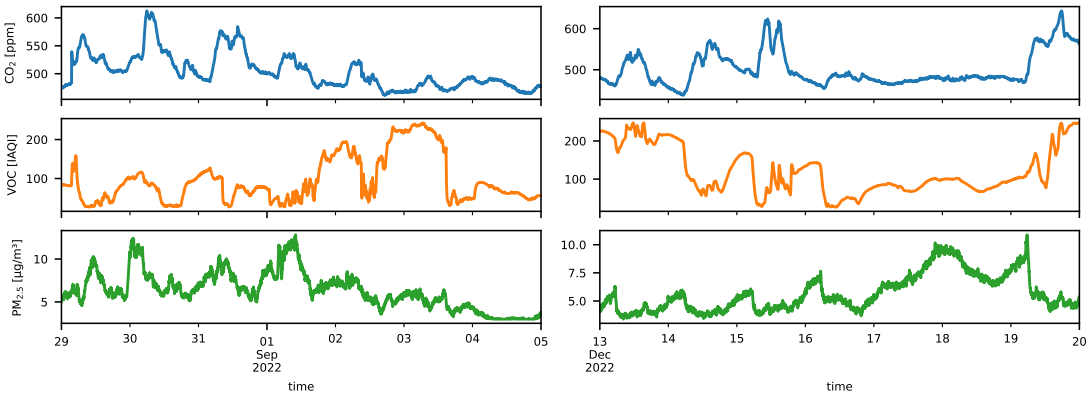


Figure 6.10: Exemplary summer (left) and winter (right) week time series plots for indoor air pollutants in Office 2

The figure shows time series plots for each pollutant for an exemplary winter and summer week. It can be seen that CO_2 follows a strong daily frequency for weekdays determined by occupancy levels. This can be observed both in the winter and summer weeks.
In the case of $PM_{2.5}$, a daily pattern is only identifiable for the winter case. Peaks in $PM_{2.5}$ concentration are situated during nighttime hours.
VOCs show no discernible pattern for both winter and summer cases. VOC shows sharp pollution peaks for both winter and summer cases during the daytime.

Seasonality The figure shows the seasonality boxplots for the three indoor air pollutants. For CO_2, no seasonality is discernible. VOCs show higher concentrations and higher volatility towards high concentrations during winter months. In the case of $PM_{2.5}$, concentrations, and variability are lower for months 1 - 4, and months 9 - 12 show higher variability and higher mean concentrations. The month of August is an outlier with significantly higher whiskers and mean values.

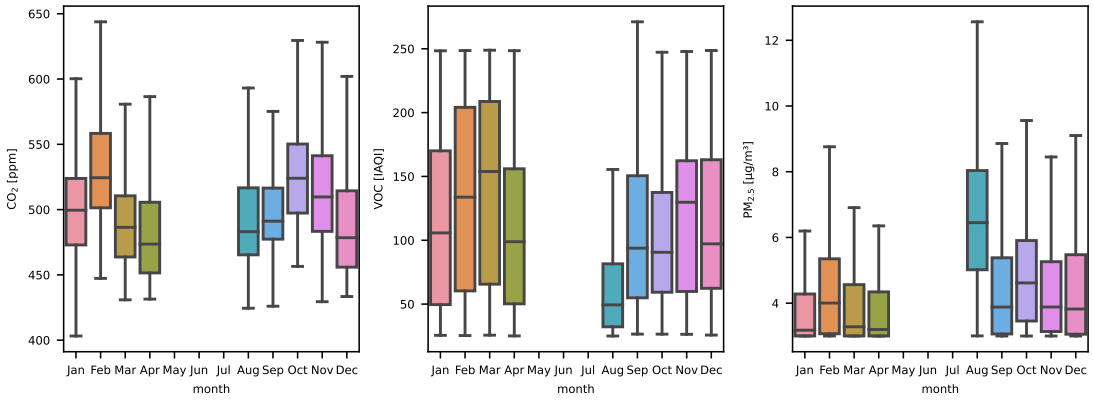


Figure 6.11: Monthly seasonality boxplots for indoor air pollutants in Office 2

6.2.3 Pollution levels

In the following, the pollution levels are compared to WHO limits, and the number of overshoot hours is calculated for each pollutant. The figure shows the number of hours of

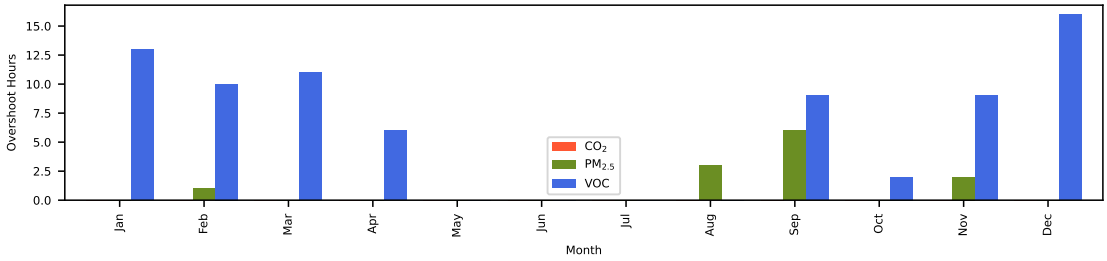


Figure 6.12: Monthly overshoot hours for indoor air pollutants during presence in Office 2

presence during which a pollutant exceeds the limits set by the WHO. In the case of CO_2, no overruns are recorded. Limited overruns for $PM_{2.5}$ concentrations were examined in February, August, September, and November. Systematical overshoots in VOC pollution were present in nearly each recorded month except August. Generally higher number of overshoot hours for VOC were identified during winter months.

6.2.4 Correlation analysis

Furthermore, the correlations between the different measurements are analyzed. These include the correlation between indoor pollutant measurements and BMS data, the correlation between indoor pollutant measurements and outdoor measurements, metadata and indoor pollutant measurements, and a time-related analysis.

Indoor - Outdoor correlation The correlation matrix shows the Pearson correlation between different outdoor measurements and indoor air pollutants. In the case of CO_2,

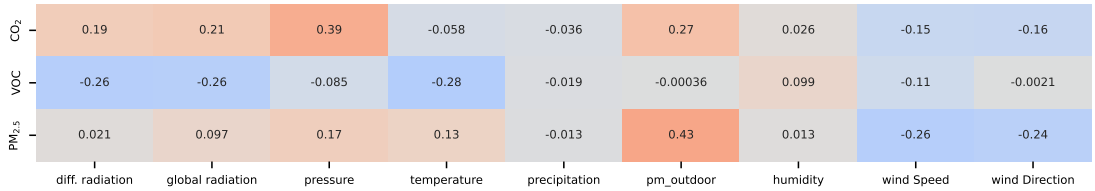


Figure 6.13: Correlation matrix: indoor air pollutants - outdoor measurements in Office 2

a moderate positive correlation was identified to outdoor air pressure. Also, weak positive correlations could be identified for radiation and outdoor $PM_{2.5}$ concentration and weak negative correlations for wind speed and direction. In the case of VOC, weak negative correlations were found for radiation, temperature, and wind speed. $PM_{2.5}$ has a moderate positive correlation with outdoor $PM_{2.5}$ concentration. Furthermore, weak positive correlations towards pressure and temperature were identified, and weak negative correlations towards wind speed and direction.

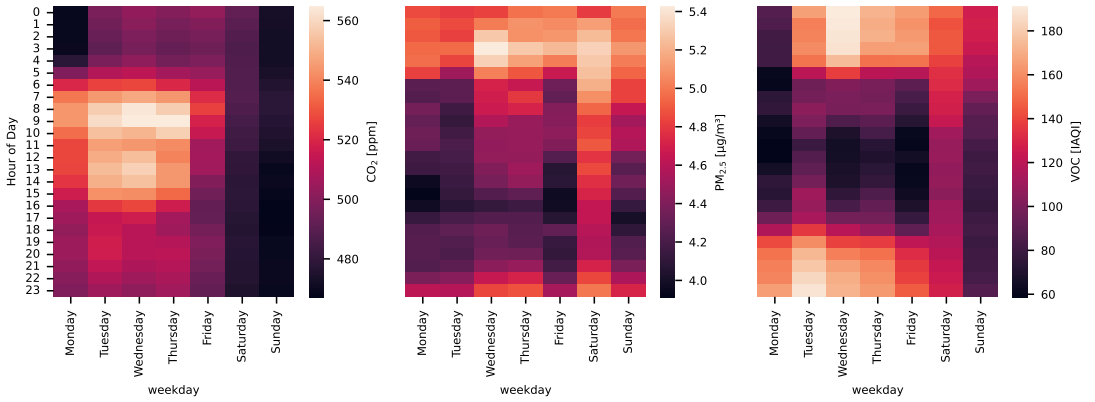


Figure 6.14: Time dependant indoor pollutant concentration in Office 2

Presence - Time correlation The heatmap shows the hour on the y-axis and the weekday on the x-axis. The color of the cells is based on the mean pollutant concentration during the specific hour and weekday for the whole measurement period. The left plot visualizes the dependence of CO_2 on time and weekdays. A strong correlation is visible, with maximum CO_2 concentration during weekday morning and early afternoon hours, with low concentration during night and weekends. Furthermore, Friday and, to a reduced degree, Monday show a reduced CO_2 concentration due to the prevailing home office.
The middle plot shows the concentration of $PM_{2.5}$ regarding time and weekdays. A reduced $PM_{2.5}$ concentration is examined between 5 am and 8 pm on weekdays. During nights and weekends, $PM_{2.5}$ concentration gradually builds up (HVAC inactivity period). Furthermore, an increased $PM_{2.5}$ concentration can be observed during the daytime on Wednesdays and Thursdays, most likely related to higher occupancy.

The third plot on the right shows the time and weekday dependence of VOC concentration. Again VOC concentration is strongly influenced by the ventilation schedule and decreases between 5 am and 8 pm during weekdays. Analog to $PM_{2.5}$, VOC tends to build up during night and weekend periods, where no ventilation is present to disperse the pollutants.

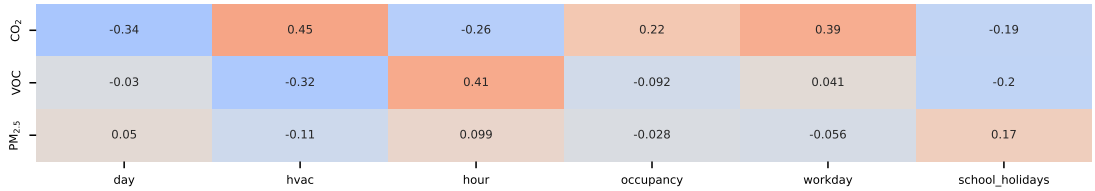


Figure 6.15: Correlation matrix: indoor air pollutants - metadata in Office 2

Metadata correlation The correlation matrix shows the Pearson correlation coefficients for indoor air pollutants and metadata records. In the case of CO_2, a moderate positive correlation can be determined for the HVAC operation, for the tag indicating whether a day is a workday or not, and a moderate negative correlation regarding the day of the week. Furthermore, a weak negative correlation was identified for the hour and school holidays and a weak positive correlation for room occupancy. In the case of VOC, a moderate positive correlation is identified for the hour and a moderate negative correlation for the HVAC operation. Furthermore, a weak negative correlation was identified for school holidays. $PM_{2.5}$ showed a weak positive correlation for school holidays and a weak negative correlation to the HVAC operation schedule; apart from this, no correlations were identified for $PM_{2.5}$ for day of week, hour, occupancy, or workday.

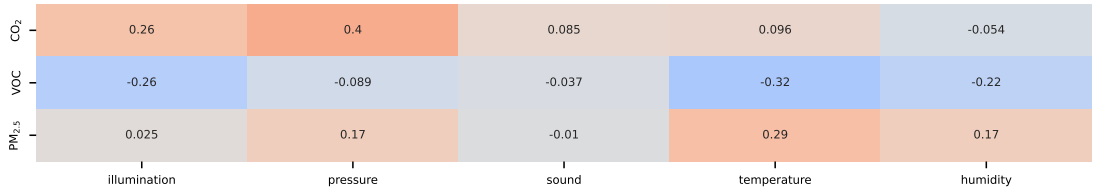


Figure 6.16: Correlation matrix: indoor air pollutants - indoor measurements in Office 2

BMS data correlation The correlation matrix shows the Pearson correlation coefficients for indoor air pollutants and indoor BMS data. CO_2 shows a moderate positive correlation to air pressure and a weak positive correlation towards illumination while showing little to no correlation towards pressure, temperature, sound, and humidity.
VOC shows a moderate negative correlation to temperature and weak negative correlations to temperature, humidity, and illumination.
In the case of $PM_{2.5}$, a weak positive correlation is determined for temperature, humidity, and air pressure.

6.3 Office 3

Data was collected in Office 3 from August 2022 to April 2023. Data was gathered from a single measurement node with a measurement interval of 60 seconds. All in all, resulting in 576 thousand measured time steps for 74 individual data points spanning indoor and outdoor measurements and metadata. All in all, 42 million data points were collected in Office 3.

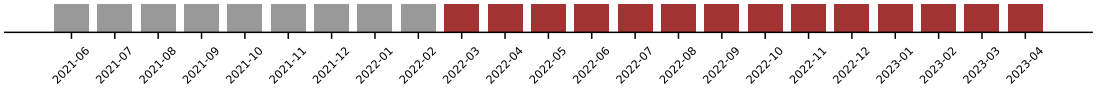


Figure 6.17: Data collection period in Office 3 (red) vs total collection period (grey)

6.3.1 Statistical analysis

Key statistical values were calculated for the indoor measurements, including indoor air pollutant measurements (CO_2, VOC, $PM_{2.5}$) and BMS data (illumination, pressure, sound, temperature, humidity). For each measurement, mean, standard deviation (std), minimum (min), 25 percentil (25%), 50 percentil (50%), 75 percentil (75%) and maximum (max) were calculated. The table summarizes the statistical values for indoor

Table 6.3: Statistical analysis for indoor measurements in Office 3

	CO_2	**VOC**	**$PM_{2.5}$**	**illu.**	**press.**	**sound**	**temp.**	**hum.**
mean	535.9	174.0	1.5	771.1	950.9	64.2	22.5	32.2
std	58.9	72.9	1.0	1358.0	8.1	1.0	1.8	8.7
min	400.0	6.0	0.0	0.0	918.5	64.0	20.0	11.5
25%	497.0	118.0	1.0	0.0	946.6	64.0	21.4	25.0
50%	524.0	203.0	1.0	13.0	951.7	64.0	22.0	31.2
75%	566.0	236.0	2.0	1094.0	955.9	64.0	23.0	38.8
max	861.0	279.0	11.0	10000.0	970.2	91.0	36.3	57.0

measurements in Office 2. Several assessments can be made based on the data. **CO_2:** The data shows a mean of CO_2 only 135 ppm over outdoor concentration and a small standard deviation of 58.9 ppm. Furthermore, percentile data shows that up to the 75. percentile, all data lies below 600 ppm, and the maximum concentration only reaches 861 ppm.
$PM_{2.5}$: $PM_{2.5}$ concentration has a mean of 1.5 $\mu g/m^3$. Variability in $PM_{2.5}$ concentration is small, and peaks only reach up to 11 $\mu g/m^3$, which is acceptable for short periods.
VOC: VOC shows higher variability than the other two air pollutants and measurements above 50. percentile exceed VOC IAQI thresholds. Thus, the VOC pollutant threshold is passed about half of the measured time.
Temperature and humidity: Data for temperature and humidity show a very static

indoor environment with minimal variations. Percentiles 25 - 75 for temperature and humidity are close to each other, and the standard deviation is low in both cases. However, compared to Office 1 and Office 2 in the same building, office 3 shows higher variability in temperature and humidity. This can be traced back to the relatively higher outdoor wall and window area.

6.3.2 Time-series analysis

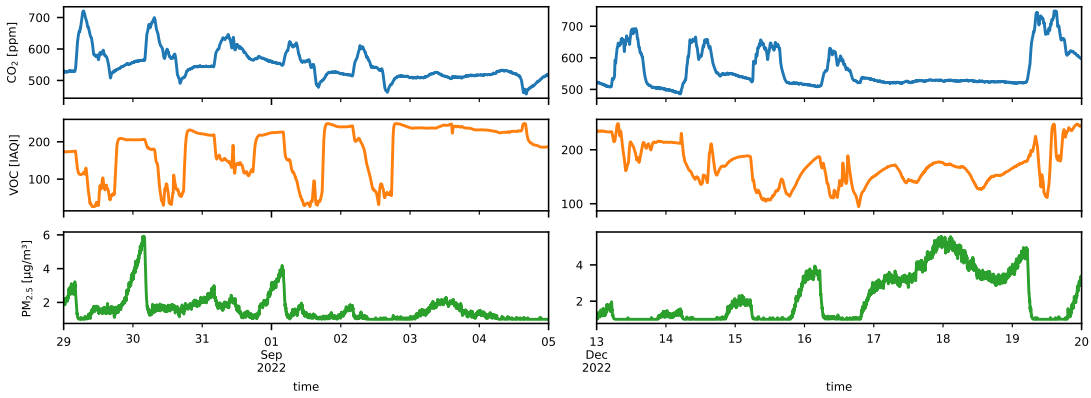


Figure 6.18: Exemplary summer (left) and winter (right) week time series plots for indoor air pollutants in Office 3

The figure shows time series plots for each pollutant for an exemplary winter and summer week. It can be seen that CO_2 follows a strong daily frequency for weekdays determined by occupancy and levels out during weekends. This can be observed both in the winter and summer weeks. In the case of $PM_{2.5}$, a daily pattern is only identifiable for the winter case. Peaks in $PM_{2.5}$ concentration are situated during nighttime hours. Furthermore, a build-up of $PM_{2.5}$ can be examined during weekend periods. VOC concentration shows a diurnal pattern more pronounced during the summer week. The activation of the ventilation system dominates the VOC concentration. If activated, VOC levels sink immediately to a low level; during nighttime and weekend hours, VOC concentration plateaus at a high concentration level.

Seasonality The figure shows the seasonality boxplots for the three indoor air pollutants. For CO_2, no seasonality is discernible.
VOCs show higher variability and mean pollutant concentrations during months 1 -3. Months 9 -12 have similar mean concentrations, however, with reduced variability.
In the case of $PM_{2.5}$, concentrations, and variability are lower for months 1 - 4 and 9, and months 10 - 12 show higher variability and higher mean concentrations. The month of August is an outlier with significantly higher whiskers and mean values. The difference for the month of August can be explained by differing occupancy patterns and possible other HVAC control due to the summer holidays.

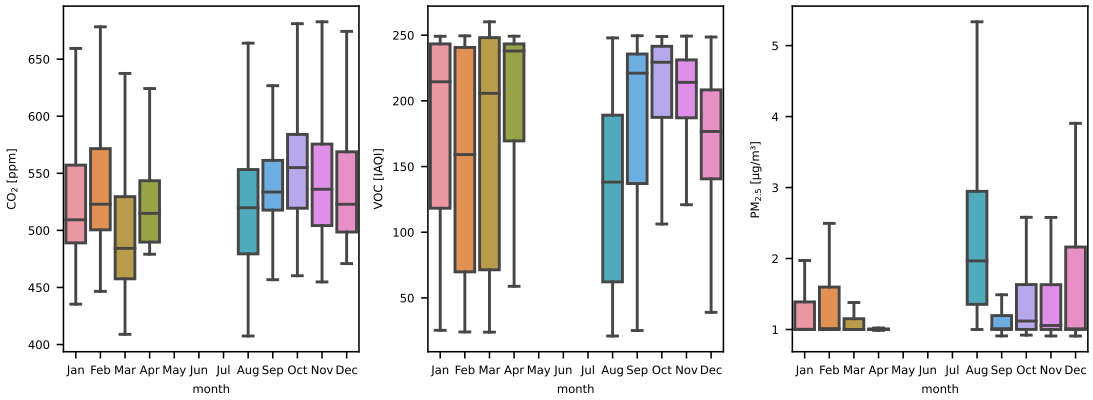


Figure 6.19: Monthly seasonality boxplots for indoor air pollutants in Office 3

6.3.3 Pollution levels

In the following, the pollution levels are compared to WHO limits, and the number of overshoot hours is calculated for each pollutant. The figure shows the number of hours

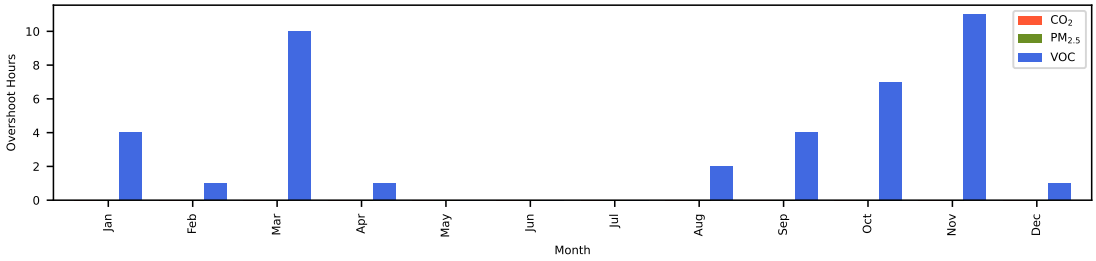


Figure 6.20: Monthly overshoot hours for indoor air pollutants during presence in Office 3

of presence during which a pollutant exceeds the limits set by the WHO for each month. No overruns were detected for CO_2 or $PM_{2.5}$. Systematical overshoots in VOC pollution were present in each recorded month. VOC overshoot hours build-up from August to November; apart from this, no pattern is discernible.

6.3.4 Correlation analysis

Furthermore, the correlations between the different measurements are analyzed. These include the correlation between indoor pollutant measurements and BMS data, the correlation between indoor pollutant measurements and outdoor measurements, metadata and indoor pollutant measurements, and a time-related analysis.

Indoor - Outdoor correlation The correlation matrix shows the Pearson correlation between different outdoor measurements and indoor air pollutants. In the case of CO_2, weak positive correlations could be identified for radiation, air pressure, and outdoor

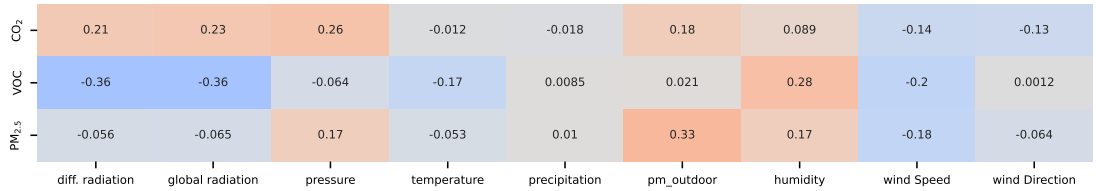


Figure 6.21: Correlation matrix: indoor air pollutants - outdoor measurements in Office 3

$PM_{2.5}$ concentration, as well as weak negative correlations for wind speed and direction. In the case of VOC, a moderate negative correlations were found for radiation and weak negative correlations for temperature and wind speed. Furthermore, a weak positive correlation was identified for humidity.
$PM_{2.5}$ has a moderate positive correlation with outdoor $PM_{2.5}$ concentration. Furthermore, weak positive correlations towards pressure and humidity and weak negative correlations towards wind speed were identified.

Presence - Time correlation The following section examines the dependence between time, weekday, and pollution concentration. The heatmap shows the hour on the y-axis

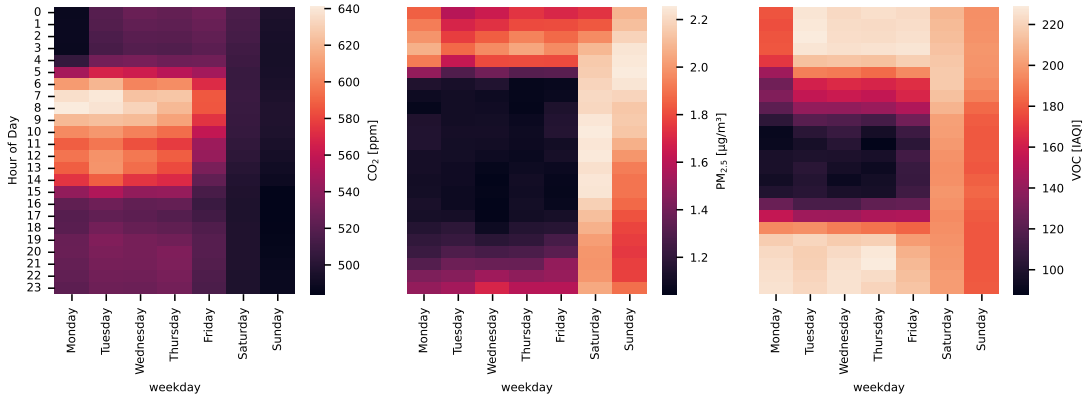


Figure 6.22: Time dependant indoor pollutant concentration in Office 3

and the weekday on the x-axis. The color of the cells is based on the mean pollutant concentration during the specific hour and weekday for the whole measurement period. The left plot visualizes the dependence of CO_2 on time and weekdays. A strong correlation is visible, with maximum CO_2 concentration during weekday morning and early afternoon hours, with low concentration during night and weekends. Furthermore, Friday shows a reduced CO_2 concentration compared to other weekdays.
The middle plot shows the concentration of $PM_{2.5}$ regarding time and weekdays. A reduced $PM_{2.5}$ concentration is examined between 5 am and 8 pm on weekdays. Compared to other rooms in the building, no daytime peaks can be observed due to the absence of operable windows. During nights and weekends, $PM_{2.5}$ concentration can gradually

build up.
The third plot on the right shows the time and weekday dependence of VOC concentration. Again VOC concentration is strongly influenced by the ventilation schedule. Analog to $PM_{2.5}$, VOC tends to build up during night and weekend periods, where no ventilation is present to disperse the pollutants.

Metadata correlation The following section explores the correlation between collected metadata and indoor air pollutants. The correlation matrix above shows the Pearson

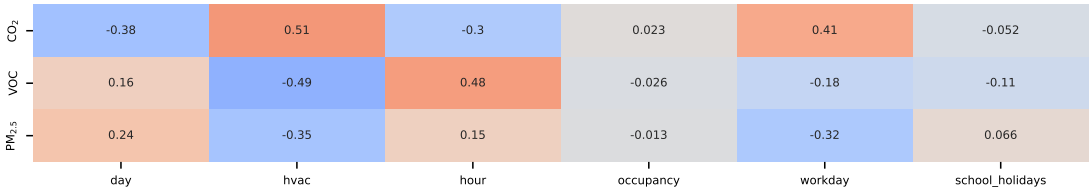


Figure 6.23: Correlation matrix: indoor air pollutants - metadata in Office 3

correlation coefficients for indoor air pollutants and metadata records. In the case of CO_2, a moderate positive correlation can be determined for the HVAC operation and for the tag indicating whether a day is a workday. Furthermore, a moderate negative correlation was identified for the hour (cyclically encoded) and the day of the week.
In the case of VOC, a moderate positive correlation is identified for the hour and a moderate negative correlation for the HVAC operation. Furthermore, a weak positive correlation is identified for the day of the week and a weak negative correlation for school holidays.
$PM_{2.5}$ shows a moderate negative correlation between HVAC operation and workdays. A weak positive correlation between the hour and day of the week can also be identified.

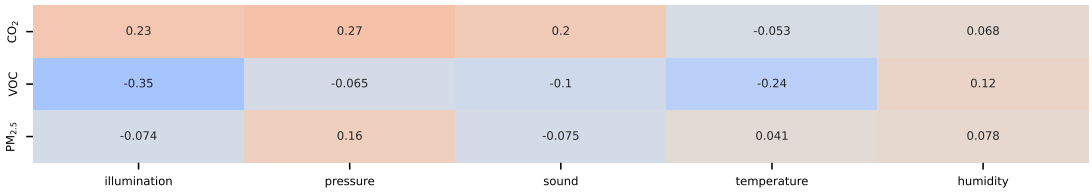


Figure 6.24: Correlation matrix: indoor air pollutants - indoor measurements in Office 3

BMS data correlation The correlation matrix shows the Pearson correlation coefficients for indoor air pollutants and indoor BMS data. CO_2 shows a weak positive correlation to air pressure, sound, and illumination while showing little to no correlation towards temperature and humidity. VOC shows a moderate negative correlation to illumination, a weak negative correlation to temperature, and a weak positive correlation to humidity. In the case of $PM_{2.5}$, a weak positive correlation can only be determined for air pressure.

6.4 Office 4

Data was collected in Office 4 from August 2022 to April 2023. Data was gathered from a single measurement node with a measurement interval of 10 seconds. All in all, resulting in 2.1 million measured time steps for 61 individual data points spanning indoor and outdoor measurements and metadata. All in all, 128 million data points were collected.

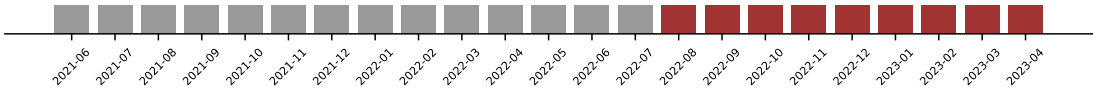


Figure 6.25: Data collection period in Office 4 (red) vs total collection period (grey)

6.4.1 Statistical analysis

Key statistical values were calculated for the indoor measurements, including indoor air pollutant measurements (CO_2, VOC, $PM_{2.5}$) and BMS data (illumination, pressure, sound, temperature, humidity). For each measurement, mean, standard deviation (std), minimum (min), 25 percentil (25%), 50 percentil (50%), 75 percentil (75%) and maximum (max) were calculated. The table summarizes the statistical values for indoor measurements in Office 4.

Table 6.4: Statistical analysis for indoor measurements in Office 4

	CO_2	**VOC**	**$PM_{2.5}$**	**illu.**	**press.**	**sound**	**temp.**	**hum.**
mean	509.3	48.0	0.6	3.5	954.1	53.7	22.0	41.8
std	44.7	37.6	0.7	6.5	6.2	6.5	1.0	4.1
min	414.8	0.0	0.0	0.0	935.8	43.5	18.8	38.1
25%	480.8	17.0	0.1	0.0	950.3	48.5	21.3	39.4
50%	498.2	42.5	0.3	1.2	955.1	51.1	21.8	40.3
75%	524.8	72.0	0.7	4.9	958.1	58.2	22.6	42.1
max	1198.5	249.5	11.8	129.4	972.4	93.1	24.7	76.7

CO_2: The data shows a mean CO_2 concentration only 109 ppm above outdoor concentration and a small standard deviation of 44.7 ppm. Furthermore, percentile data shows that up to the 50. percentile, all data lies below 500 ppm, and the maximum concentration reaches 1198 ppm.
$PM_{2.5}$: $PM_{2.5}$ concentration is very low with a mean of 0.6 $\mu g/m^3$. Furthermore, variability in $PM_{2.5}$ concentration is also small, and peaks only reach up to 11 $\mu g/m^3$, which is acceptable for short periods.
VOC: VOC shows slightly higher variability than the other two air pollutants, even though measurements are up to 75. percentile and above are below health-related thresholds. However, short-term overshoots exist with a recorded maximum value of 249 IAQI.

6.4.2 Time-series analysis

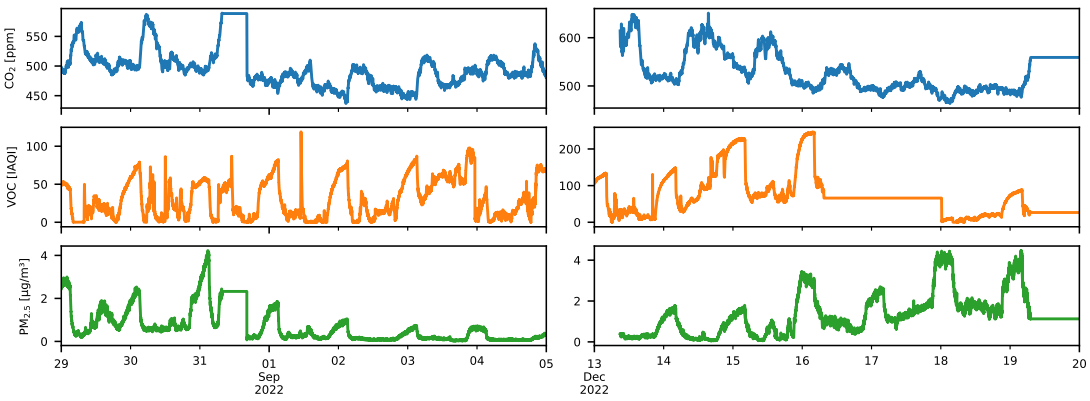


Figure 6.26: Exemplary summer (left) and winter (right) week time series plots for indoor air pollutants in Office 4

The figure shows time series plots for each pollutant for an exemplary winter and summer week. It can be seen that CO_2 follows a strong diurnal pattern for weekdays determined by occupancy and levels out during weekends. This can be observed both in the winter and summer weeks.
In the case of $PM_{2.5}$, a daily pattern is only identifiable for the winter case. Peaks in $PM_{2.5}$ concentration are situated during nighttime hours. Furthermore, a build-up of $PM_{2.5}$ can be examined during weekend periods.
VOCs show less of a pattern for both winter and summer cases. In the winter, an increase during nighttime hours can be examined, which rapidly decreases when the HVAC is turned on at 5:15. Apart from this, VOC shows sharp pollution peaks for both winter and summer cases during the daytime.

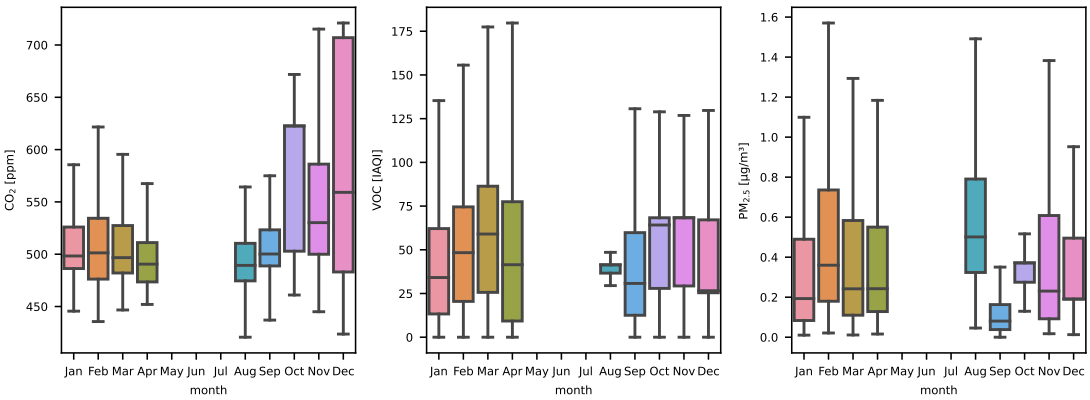


Figure 6.27: Monthly seasonality boxplots for indoor air pollutants in Office 4

Seasonality The figure shows the seasonality boxplots for the three indoor air pollutants. For CO_2, no seasonality is discernible. However, an increase in variability and mean pollutant concentration can be seen for October - December.
VOCs show higher variability and mean pollutant concentrations during January - April. Furthermore, an increased mean concentration can be examined in August.
In the case of $PM_{2.5}$ concentrations, no seasonality pattern is discernible.

6.4.3 Pollution levels

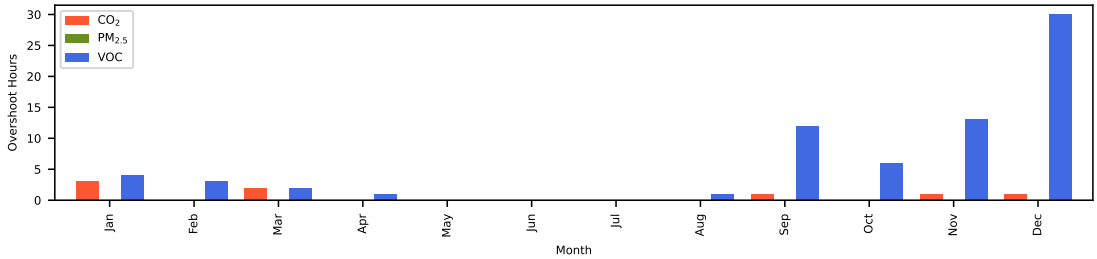


Figure 6.28: Monthly overshoot hours for indoor air pollutants during presence in Office 4

The figure shows the number of hours of presence during which a pollutant exceeds the limits set by the WHO. No overruns were detected for $PM_{2.5}$. Minimal hours of overshoot of CO_2 are present. Systematical overshoots in VOC pollution were present in each recorded month, with a sharp rise from October to December.

6.4.4 Correlation analysis

Furthermore, the correlations between the different measurements are analyzed. These include the correlation between indoor pollutant measurements and BMS data, the correlation between indoor pollutant measurements and outdoor measurements, metadata and indoor pollutant measurements, and a time-related analysis.

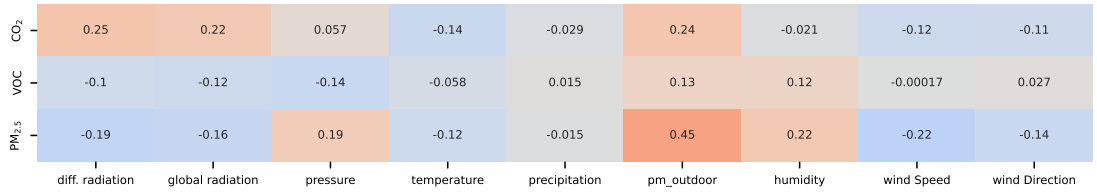


Figure 6.29: Correlation matrix: indoor air pollutants - outdoor measurements in Office 4

Indoor - Outdoor correlation The correlation matrix shows the Pearson correlation between outdoor measurements and indoor air pollutants. In the case of CO_2, weak positive correlations could be identified for radiation and outdoor $PM_{2.5}$ concentration

and weak negative correlations for temperature, wind speed, and direction.
In the case of VOC, weak negative correlations were examined for air pressure and global radiation. Furthermore, a weak positive correlation was identified between outdoor $PM_{2.5}$ and humidity.
$PM_{2.5}$ has a moderate positive correlation with outdoor $PM_{2.5}$ concentration. Furthermore, weak positive correlations towards pressure and humidity were identified, and weak negative correlations towards wind speed, wind direction, radiation, and temperature.

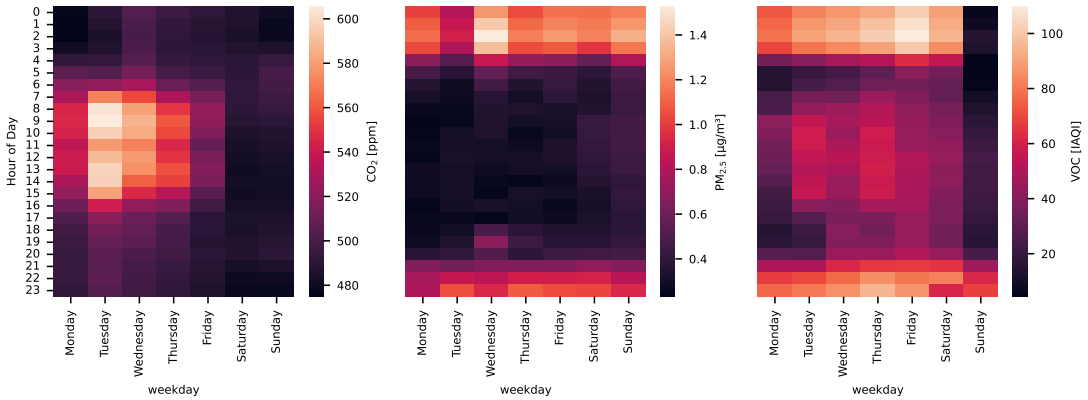


Figure 6.30: Time dependant indoor pollutant concentration in Office 4

Presence - Time correlation The heatmap shows the hour on the y-axis and the weekday on the x-axis. The color of the cells is based on the mean pollutant concentration during the specific hour and weekday for the whole measurement period. The left plot visualizes the dependence of CO_2 on time and weekdays. A strong correlation is visible, with maximum CO_2 concentration during weekday morning and early afternoon hours, with low concentration during night and weekends. Furthermore, Friday and, to a reduced degree, Monday show a reduced CO_2 concentration.
The middle plot shows the concentration of $PM_{2.5}$ regarding time and weekdays. A reduced $PM_{2.5}$ concentration is seen between 5 am and 8 pm on weekdays, induced by ventilation. During nights and weekends, $PM_{2.5}$ concentration gradually builds up.
The third plot on the right shows the time and weekday dependence of VOC concentration. Again VOC concentration is strongly influenced by the ventilation schedule and decreases between 5 am and 8 pm during weekdays. However, VOC concentration is elevated, especially during days with higher occupancy (Tuesday, Wednesday, and Thursday). Analog to $PM_{2.5}$, VOC tends to build up during night and weekend periods, where no ventilation is present to disperse the pollutants.

Metadata correlation The following section explores the correlation between collected metadata and indoor air pollutants. The correlation matrix above shows the Pearson correlation coefficients for indoor air pollutants and metadata records. In the case of CO_2, a moderate positive correlation can be determined for the HVAC operation and

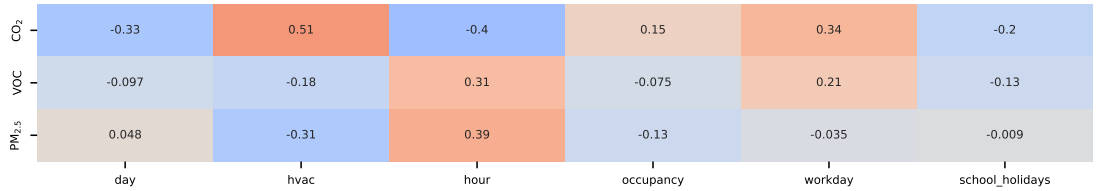


Figure 6.31: Correlation matrix: indoor air pollutants - metadata in Office 4

for the tag indicating whether a day is a workday. Furthermore, a moderate negative correlation was identified for the hour (cyclically encoded) and the day of the week. A weak positive correlation is found for occupancy, and a weak negative correlation during school holidays.
In the case of VOC, a moderate positive correlation is identified for the hour. Furthermore, a weak positive correlation is identified for workdays and a weak negative correlation for school holidays and HVAC operation.
$PM_{2.5}$ shows a moderate negative correlation towards HVAC operation and a moderate positive correlation to the hour of the day. Additionally, a weak negative correlation for occupancy can be identified.

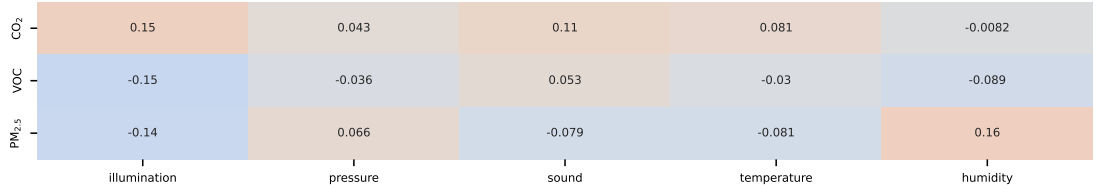


Figure 6.32: Correlation matrix: indoor air pollutants - indoor measurements in Office 4

BMS data correlation The correlation matrix shows the Pearson correlation coefficients for indoor air pollutants and indoor BMS data. CO_2 shows a weak positive correlation to air pressure and illumination while showing little to no correlation towards temperature and humidity. VOC shows a weak negative correlation to illumination. In the case of $PM_{2.5}$, a weak positive correlation can be determined for humidity and a weak negative correlation for illumination.

6.5 Classroom

Data was collected in the Classroom from March 2022 to April 2023. Data was gathered from a single measurement node with a measurement interval of 60 seconds. All in all, resulting in 576 thousand measured time steps for 95 individual data points spanning indoor and outdoor measurements and metadata. All in all, 55 million data points were collected.

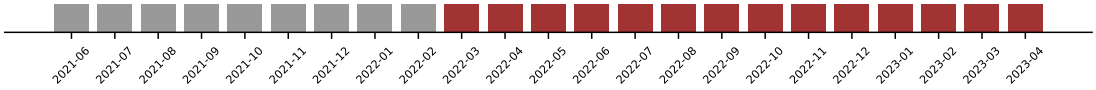


Figure 6.33: Data collection period in the classroom (red) vs total collection period (grey)

6.5.1 Statistical analysis

Key statistical values were calculated for the indoor measurements, including indoor air pollutant measurements (CO_2, VOC, $PM_{2.5}$) and BMS data (illumination, pressure, sound, temperature, humidity). For each measurement, mean, standard deviation (std), minimum (min), 25 percentil (25%), 50 percentil (50%), 75 percentil (75%) and maximum (max) were calculated. The table summarizes the statistical values for indoor

Table 6.5: Statistical analysis for indoor measurements in the classroom

	CO_2	**VOC**	**$PM_{2.5}$**	**illu.**	**press.**	**sound**	**temp.**	**hum.**
mean	886.5	127.9	6.3	75.3	957.8	65.0	22.0	38.4
std	551.7	62.1	4.5	132.5	8.3	4.0	4.1	5.7
min	400.0	0.0	2.0	0.0	472.9	64.0	-4.1	14.0
25%	531.0	72.0	4.0	0.5	954.2	64.0	18.9	36.3
50%	633.6	130.0	5.0	2.0	958.3	64.0	21.2	39.3
75%	1016.5	177.0	7.0	100.5	962.3	64.0	25.2	42.0
max	4713.0	413.0	91.0	2845.5	982.2	99.0	32.6	64.0

measurements in the Classroom. Several assessments can be made based on the data.
CO_2: The data shows that CO_2 has a mean of 486 ppm above outdoor concentration with high standard deviation of 551.7 ppm, indicating a high variability and high CO_2 concentrations in the Classroom. 25% of the measurements surpass the 1000 ppm indoor concentration threshold, and a maximum of 4713 ppm indicates extremely high concentrations during peak times. Compared to the office rooms Office 1 - Office 4, the Classroom has significantly higher mean and maximum concentrations and variability of CO_2.
$PM_{2.5}$: $PM_{2.5}$ concentration has a mean of 6.3 $\mu g/m^3$ and a standard deviation of 4.5 $\mu g/m^3$. Even though measurements up to 75. percentile is below the WHO threshold for $PM_{2.5}$ concentrations, a maximum of 91 $\mu g/m^3$ shows high pollutant events with

health-relevant concentrations. Compared to the office rooms, a significantly higher mean value of $PM_{2.5}$ concentration is examined in the Classroom, as well as a larger variability.
VOC: VOC shows less variability than the other two air pollutants and is similar to the situation in Office 2. Measurements up to 75. percentile is below health-related thresholds. However, short-term overshoots with a recorded maximum value of 413 IAQI indicate short-term high pollution events. Compared to the office rooms, variability in VOC is similar to that in Office 2 and Office 3; compared to the other indoor air pollutants, less difference exists between Office and Classroom typologies for VOC.

6.5.2 Time-series analysis

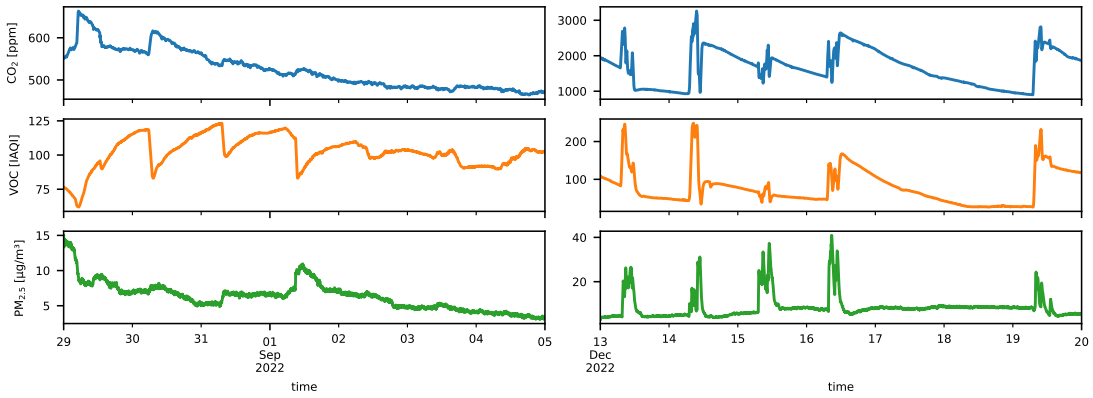


Figure 6.34: Exemplary summer (left) and winter (right) week time series plots for indoor air pollutants in the classroom

The figure shows time series plots for each pollutant for an exemplary winter and summer week. It can be seen that occupancy patterns highly influence CO_2. Due to high occupant density in the room, CO_2 concentrations rise sharply up to peaks of 3000 ppm during regular Classroom use hours, followed by a gradual reduction of concentration based on infiltration and ventilation. This effect is less pronounced in the summer week due to natural ventilation.
VOCs are strongly correlated to occupancy and show sharp peaks during occupied hours; VOC concentrations decay with time after occupation ceases.
In the case of $PM_{2.5}$, a strong daily pattern can also be observed for the winter week and with reduced intensity in the summer week. $PM_{2.5}$ sharply rises during occupancy and falls back to a baseline value after occupancy. In summer weeks, this effect is intensified by pollutant peaks induced by external pollution and natural ventilation.

Seasonality The figure shows the seasonality boxplots for the three indoor air pollutants for the Classroom. For CO_2, a strong seasonality is discernible, with significantly increased pollution and variability during the winter months of November - march. During April - October, natural ventilation leads to lower mean concentration values and less variability.

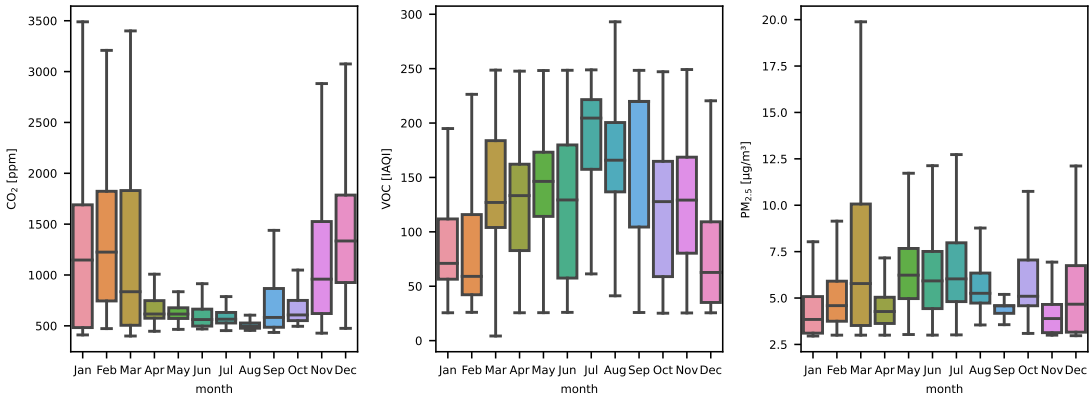


Figure 6.35: Monthly seasonality boxplots for indoor air pollutants in the classroom

VOC concentrations show moderate effects of seasonality, with lower concentrations in the winter months of December to February.
In the case of $PM_{2.5}$, no seasonality pattern is discernible; however, it is noteworthy that $PM_{2.5}$ concentrations in the march are exposed to significantly higher variability.

6.5.3 Pollution levels

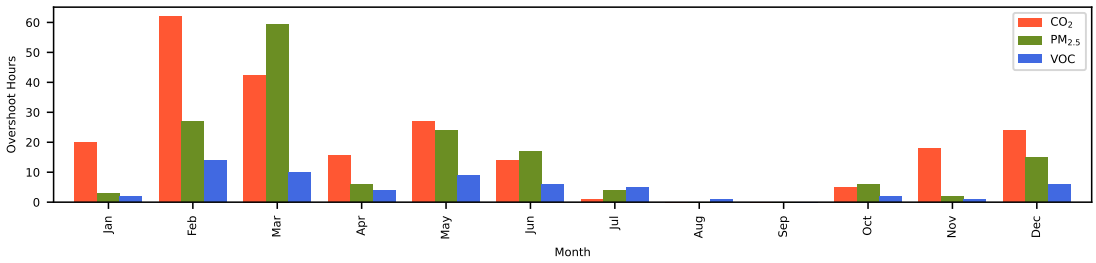


Figure 6.36: Monthly overshoot hours for indoor air pollutants during presence in the classroom

The figure shows the number of hours during which pollutants exceed the limits set by the WHO. Overruns were detected for all three pollutants, however, in varying severity. The lowest overshoots were identified for VOCs, which are relatively evenly spread throughout the year, with the exceptions of August - November, which have very low or no overshoots.
$PM_{2.5}$ exceeds the thresholds regularly, with the highest exceedances for February , March, and May. Generally, lower or no overshoots are seen during the summer months, July to September.
In the case of CO_2, many overshoot hours are present, especially in the winter months of November - March. In the transitional period of April - June and October, low overshoots were identified, and no overshoots were seen for the summer months of July - September.

6.5.4 Correlation analysis

Furthermore, the correlations between the different measurements are analyzed. These include the correlation between indoor pollutant measurements and BMS data, the correlation between indoor pollutant measurements and outdoor measurements, metadata and indoor pollutant measurements, and a time-related analysis.

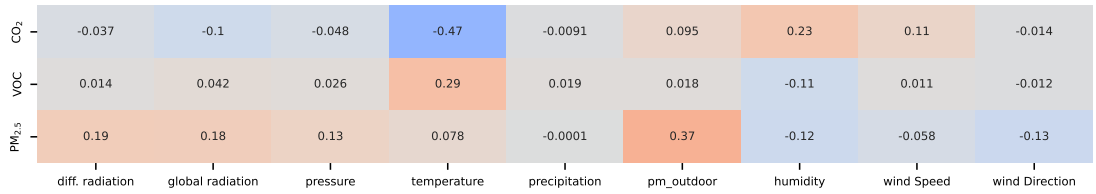


Figure 6.37: Correlation matrix: indoor air pollutants - outdoor measurements in the classroom

Indoor - Outdoor correlation The correlation matrix shows the Pearson correlation between outdoor measurements and indoor air pollutants. In the case of CO_2, a moderate negative correlation could be identified for outdoor temperature. Furthermore, weak positive correlations could be identified for humidity and wind speed and weak negative correlations to global radiation.
In the case of VOC, weak negative correlations were examined for outdoor humidity. Furthermore, a weak positive correlation was identified for outdoor temperature.
$PM_{2.5}$ has a moderate positive correlation with outdoor $PM_{2.5}$ concentration. Furthermore, weak positive correlations towards pressure and radiation were identified, and weak negative correlations towards wind direction and humidity.

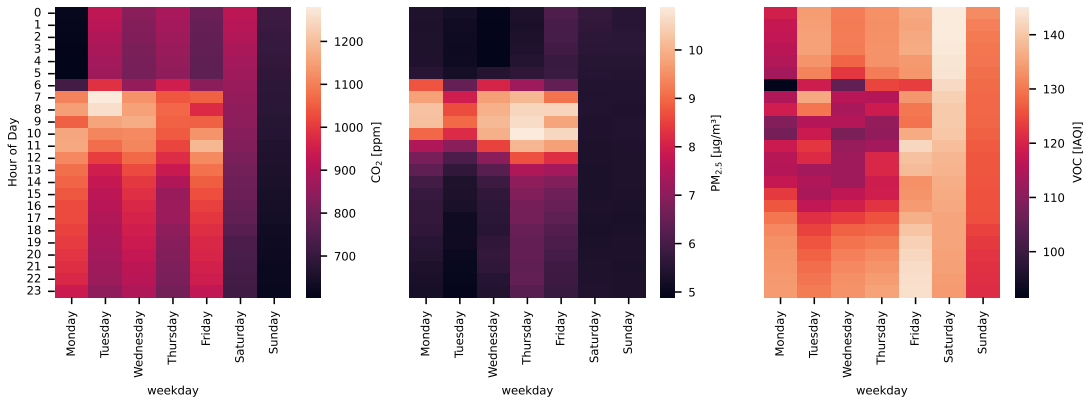


Figure 6.38: Time dependant indoor pollutant concentration in the classroom

Presence - Time correlation The heatmap shows the hour on the y-axis and the weekday on the x-axis. The left plot visualizes the dependence of CO_2 on time and

weekdays. A strong correlation is visible, with maximum CO_2 concentration during the morning hours of schooldays; after occupancy ceases in the afternoon or on weekends, CO_2 concentration decays. However, it can be seen that even though more than 12 hours of unoccupied time are between occupied periods during weekdays, CO_2 concentration regularly does not decrease to outdoor levels.
The middle plot shows the concentration of $PM_{2.5}$ regarding time and weekdays. Peaks in $PM_{2.5}$ emerge primarily during occupied times and fall back to a base concentration level during unoccupied times. The third plot on the right shows the time and weekday dependence of VOC concentration. Compared to the other pollutants, a more homogenous distribution can be identified; however, VOC concentration is generally lower during mornings and early afternoons on school days.

Metadata correlation The following section explores the correlation between collected metadata and indoor air pollutants. The correlation matrix above shows the Pearson

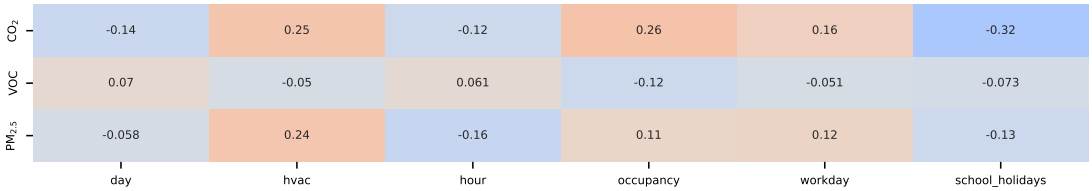


Figure 6.39: Correlation matrix: indoor air pollutants - metadata in the classroom

correlation coefficients for indoor air pollutants and metadata records. Regarding CO_2, a moderate negative correlation can be determined for school holidays. Furthermore, a weak negative correlation was identified for the hour (cyclically encoded) and the day of the week. A weak positive correlation exists for occupancy, HVAC operation, and workdays.
In the case of VOC, only a weak negative can be identified for occupancy; for all other records, little or no correlation can be examined.
$PM_{2.5}$ shows a weak negative correlation between hours and school holidays. Furthermore, a weak positive correlation was found for HVAC operation, occupancy, and workdays.

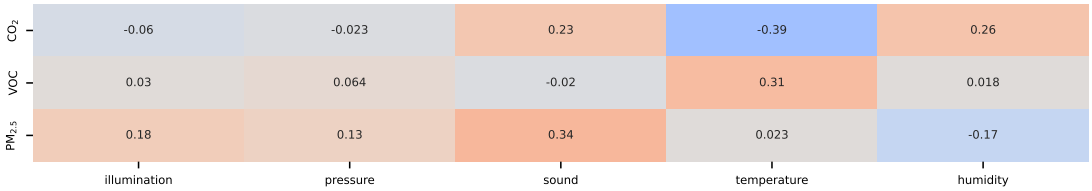


Figure 6.40: Correlation matrix: indoor air pollutants - indoor measurements in the classroom

BMS data correlation The correlation matrix shows the Pearson correlation coefficients for indoor air pollutants and indoor BMS data. CO_2 shows a moderate negative correlation

to temperature and a weak positive correlation towards humidity and sound.
In the case of VOC, only a weak positive correlation can be identified for temperature; for all other records, little or no correlation can be examined.
In the case of $PM_{2.5}$, a moderate positive correlation is identified for sound, a weak positive correlation for illumination and pressure, and a weak negative correlation for humidity.

6.6 Comparison

In this section, the measurements and results from the respective rooms are compared regarding their statistical distributions, pollution levels, and correlations.

6.6.1 Comparison of $PM_{2.5}$ concentration

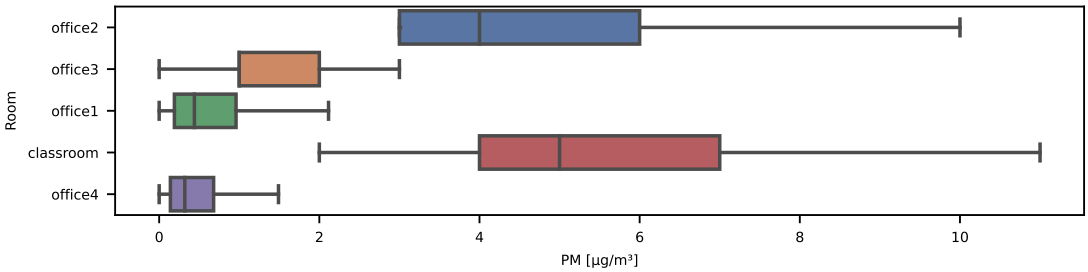


Figure 6.41: Comparison of room $PM_{2.5}$ concentration levels

The figure shows the distribution of $PM_{2.5}$ in the five rooms. Office 1 and Office 4, show very similar distribution results for $PM_{2.5}$ with overall lowest mean values and low variability. Office 3 - without operable windows and three outside facing facades has elevated pollutant levels compared to Offices 1 and 2 with increased spread. The highest $PM_{2.5}$ concentrations are identified in Office 2 and the Classroom, with the highest mean value and maximum variability present in the Classroom.

6.6.2 Comparison of CO_2 concentration

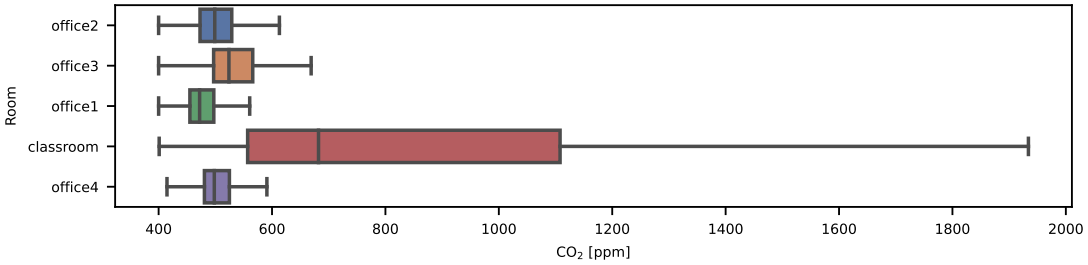


Figure 6.42: Comparison of room CO_2 concentration levels

The figure shows the distribution of CO_2 in the five rooms. The Classroom stands out, with significantly higher median and variability than the office rooms. The office rooms, on the other hand, show very similar indoor environments concerning CO_2, with low variability in CO_2 concentration and low concentrations throughout all measurements. In the group of offices, Office 3 - without operable windows - has the highest median and highest variability in terms of CO_2, due to the absence of natural ventilation.

6.6.3 Comparison of VOC concentration

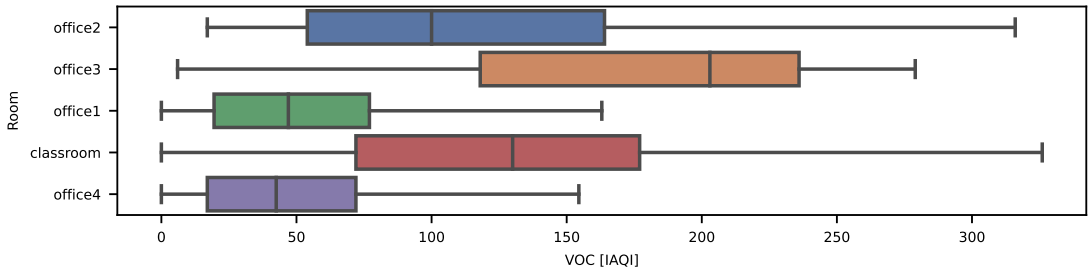


Figure 6.43: Comparison of room VOC concentration levels

The figure shows the distribution of VOC in the five rooms. Office 1 and Office 4 are very similar in concentration and distribution of VOC, achieving the overall lowest mean VOC concentrations. Office 2 and the Classroom are also similar in their level of concentration and variability. Office 3 stands out with the highest median value for VOC concentration, which most likely results from the absence of operable windows and corresponding lack of natural ventilation.

6.6.4 Comparison of pollution levels

In the following, the pollution levels throughout the different rooms are compared. The

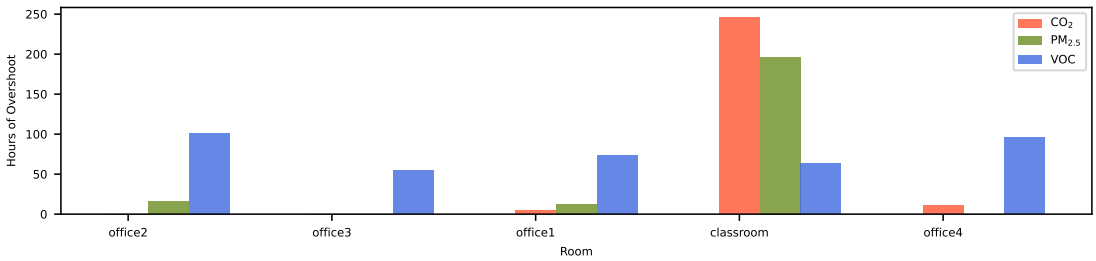


Figure 6.44: Comparison of annual pollution overshoot hours

figure shows the number of annual overshoot hours during the presence of each pollutant and room. VOC overshoots are present in every room and are within 50 and 100 hours of exceedance for every room.
However, $PM_{2.5}$ and CO_2 overshoots are mostly only present in the Classroom. Minor $PM_{2.5}$ overruns are only present in Office 2 and Office 1, and minor CO_2 overruns in Office 4. The Classroom exceeds the CO_2 concentration of 1000 ppm at more than 200 occupied hours and at close to 200 hours in the case of $PM_{2.5}$. These differences in pollution levels can be traced back to the significantly different usage patterns throughout the typologies with much higher occupant density in the Classroom. As seen in the

above section, pollution in the Classroom is primarily correlated with occupancy. In office rooms, indoor air pollutants are mainly correlated to the HVAC schedule.

6.6.5 Correlations

Correlations towards outdoor, indoor, and metadata were analyzed for all rooms. Generally, the correlations were similar for Office rooms 1, 2, and 4 (with operable windows and mechanical ventilation) and deviating for Office 3 and the Classroom.
For offices, 1 - 4, the only outdoor factor that continuously achieved a moderate correlation was outdoor $PM_{2.5}$ concentration, which positively correlated with indoor $PM_{2.5}$ concentration. This is as expected since the main source of $PM_{2.5}$ is outdoors, and the possibility of operable windows introduces pollutants into the room. Multiple metadata records correlated with indoor air pollutants, foremost, and the overall strongest correlation in the office rooms was the activity of the HVAC system, which correlated with all three indoor air pollutants but most strongly with CO_2.
Furthermore, time-related metadata for workday, hour, and day were identified to correlate with all indoor air pollutants. Indoor BMS data such as illumination, pressure, sound, temperature, and humidity achieved only weak correlations in rooms 1 and 4 but moderate correlation for pressure and temperature in Office 2 and illumination in office 3. However, it is noteworthy that the Pearson correlation coefficient only denotes the linear correlation between values, thus only capturing the direct influence between two variables.
In the case of the Classroom, the strongest outdoor correlate was outdoor temperature, followed by outdoor $PM_{2.5}$; it is surmised that outdoor temperature has an increased influence on the ventilation behavior in the Classroom since no active cooling solution is present compared to the office rooms. Compared to the office rooms, where multiple metadata records correlated with indoor air pollutants in the case of the Classroom, only school holidays achieved a moderate correlation for CO_2. In the case of BMS data only temperature and sound correlate with indoor air pollutants in the Classroom.
Overall, it can be summarized that the examined Classroom behaves significantly differently than the office rooms due to different occupancy patterns and a much higher occupant density, as well as different determinants of indoor air pollutant concentration.

7 Results virtual sensing model

Certain portions of the textual content, data, figures, and results included in this chapter have been previously published in Gabriel and Auer [2022] and Gabriel and Auer [2023].

In this section, the results of the machine-learning based virtual indoor air pollutant sensors model are presented and discussed.
Subsection 7.1 evaluates the LSTM based model. Subsection 7.2 reports the results of transferring the model to an unseen office room. Subsection 7.3 evaluates the transfer learning approach on the LSTM model transferred to Open Office 2, Office 3 and the Classroom. In the final section, the results of the LSTM model are compared to the MLP and SGD models for the training room as well as the transfer learning approach. The results are summarized in section 7.7.

7.1 LSTM Model evaluation

"Figure 7.1 displays the predictions of the trained LSTM model for the testing set in Office 1 ([blue]) and the measured truth ([red]) for each indoor air pollutant. The evaluation metrics are calculated individually for each pollutant and shown in the top-left corner of each plot. The testing was conducted for three months, from March 2022 to May 2022. A visual assessment of the time series plots reveals a high correlation between the truth and prediction. The most significant deviations between truth and prediction are identified for CO_2 predictions. In the case of CO_2, the model tends to slightly overestimate the CO_2 concentration during low-concentration periods, whereas high-concentration events show a closer fit. However, the model occasionally predicts pollutant peaks incorrectly during low-concentration periods and vice versa. In the case of CO_2 predictions, they appear to be more accurate during the second half of the testing period. [...][T]he visual assessment shows an excellent fit between prediction and truth [for $PM_{2.5}$]. All peaks are identified correctly. The time series plot demonstrates a slight underestimation of peaks and high pollution events by the prediction compared to the truth. In the case of VOC, the visual assessment of the time series plots reveals an excellent fit between prediction and truth. The model can detect all concentration peaks, even though VOC concentration is highly dynamic. However, a slight underestimation of pollutant peaks can be observed in the time series, especially during the first half of the testing period." [Gabriel and Auer, 2023] Table 7.1 summarizes each pollutant's evaluation metrics (RMSE, MAE). "[T]he model exhibits a low error for all pollutants, as demonstrated by the MAE and RMSE performance metrics. In the case of CO_2, the mean absolute error amounts to 15.4 ppm for the testing period, while a slightly

increased RMSE value of 20.2 ppm indicates that no outliers significantly impact the model's predictions. The CO_2 measurements ranged from 380 to 560 ppm during the measurement period. For [$PM_{2.5}$], the errors amount to 0.3 and 0.5 $\mu g/m^3$ for MAE and RMSE, respectively, indicating consistently low error rates[, however, with some outliers as indicated by the twice as high RMSE metric]. The measurements ranged from 0 to 13 $\mu g/m^3$ during the measurement period. In the case of [VOC], MAE and RMSE errors amount to 20.1 IAQI and 31.4 IAQI, respectively, demonstrating low error rates without major deviations. The measurements for VOC ranged from 0 to 450 IAQI." [Gabriel and Auer, 2023] Overall, the evaluation results of the virtual indoor air pollutant sensor, based

Table 7.1: Evaluation metrics LSTM virtual sensing model.

Pollutant	**MAE**	**RMSE**
CO_2	15.4	20.2
$PM_{2.5}$	0.3	0.5
VOC	20.1	31.4

on an LSTM model indicated a high correlation between actual and predicted values for $PM_{2.5}$ and VOC. "The model successfully identified all pollutant peaks during the testing period, with the only error being a slight underestimation of peak concentrations. For CO_2, a less ideal but still satisfactory prediction result was achieved. This led to minor errors and a less accurate representation of the variability in actual concentrations, resulting in erroneous predictions, such as misidentified pollutant peaks during the testing period. Nevertheless, the predictions yielded a mean absolute error within the range of measurement inaccuracies for most sensors. The performance metrics of MAE = 15.4 ppm [and] RMSE = 20.2 ppm for CO_2 showed very low errors with insignificant outliers. [...] For $PM_{2.5}$, the metrics MAE = 0.3 $\mu g/m^3$, RMSE = 0.5 $\mu g/m^3$ signified a strong prediction capability with minimal errors and an excellent fit. Similar results were observed for VOC, with MAE = 20.1 IAQI and RMSE = 31.4 IAQI [...]. Overall, the LSTM model demonstrated strong performance in predicting indoor air pollutant concentrations." [Gabriel and Auer, 2023] Therefore, the LSTM model is suitable for replacing physical indoor air pollutant sensors in this room.

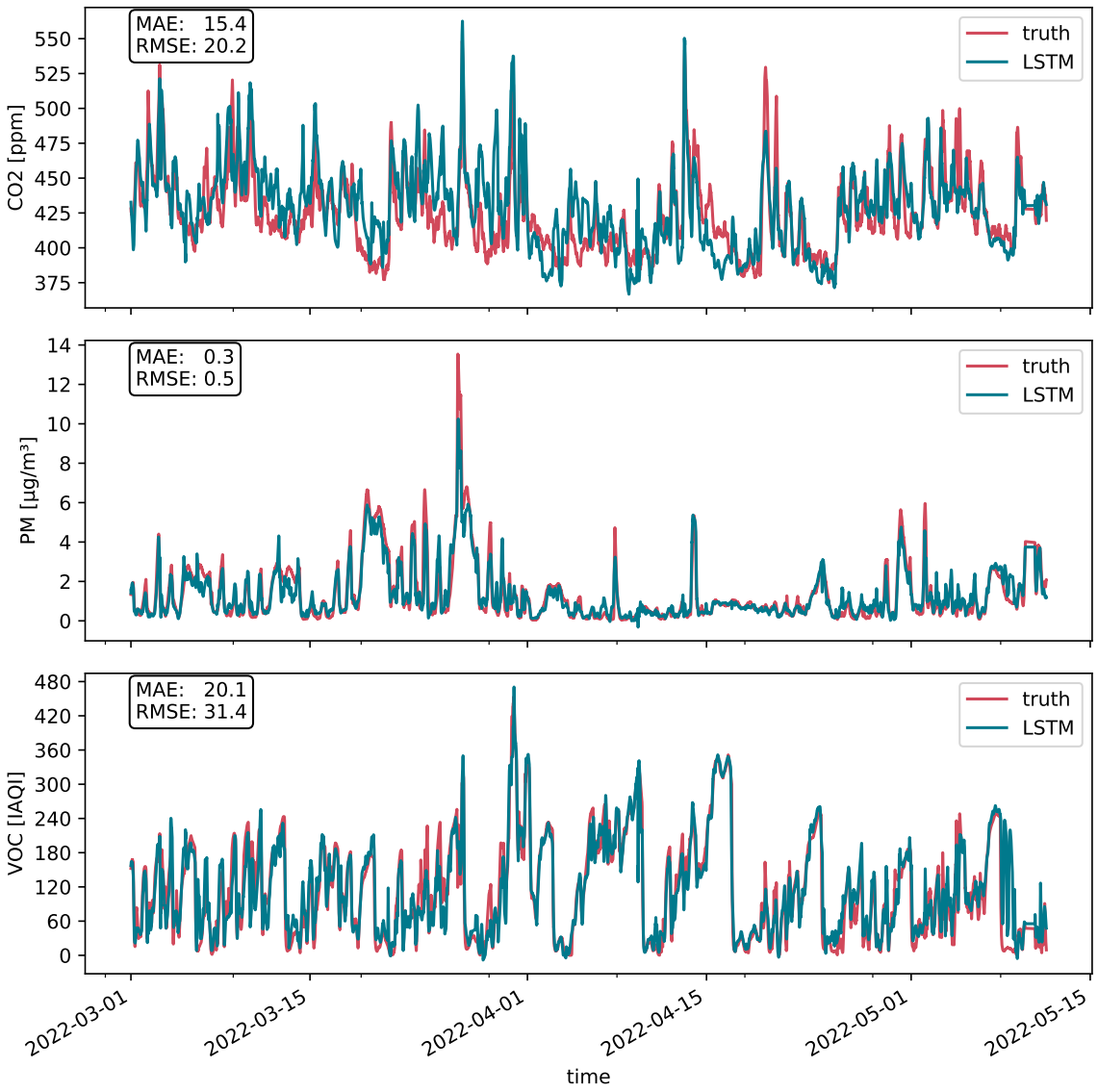


Figure 7.1: Comparison of Predicted (blue) and True (red) Time Series for Indoor Air Pollutants: VOC (bottom), $PM_{2.5}$ (middle), and CO_2 (top) [own representation]

7.2 LSTM unseen transfer evaluation

Figure 7.2 displays the predictions of the trained LSTM model (blue) for Office 4, as well as the measured actual values (red) for each indoor air pollutant. The test took place in March 2023. "Visual assessment of the time series plots revealed a correlation between actual values and predictions for all pollutants, albeit with varying degrees of fit. The highest correlation between actual and predicted values was observed for CO_2 predictions. The prediction model successfully identified pollutant peaks, albeit with underestimation. During low pollutant events, such as weekends or nights, the model results were less smooth and tended to overestimate variability in pollutant concentrations. Occasionally, the model predicted pollutant peaks under unpolluted conditions. [The visual assessment showed a general fit between the magnitudes of predicted and actual concentrations] [f]or [$PM_{2.5}$]. However, the prediction failed to detect some peaks and underestimated all others. In some cases, the prediction exhibited a phase shift, resulting in delayed identification of rising concentrations. For VOC, a visual assessment of the time series plots revealed that the model could identify some concentration peaks. However, the model frequently and erroneously detected pollutant peaks when none were present. Table 7.1 summarizes each pollutant's evaluation metrics [...]. Overall, the model exhibited very low errors for all pollutants, as evidenced by the MAE and RMSE performance metrics. For CO_2, the mean absolute error was 21.9 ppm during the testing period, while a slightly increased RMSE value of 30.4 ppm indicated no outliers affected the model's predictions. CO_2 measurements ranged from 420 ppm to 610 ppm during the measurement period. For [$PM_{2.5}$], errors amounted to 0.3 and 0.6 $\mu g/m^3$ for MAE and RMSE, respectively, indicating consistently low error rates without outliers. Measurements ranged from 0 to 4 $\mu g/m^3$ during the measurement period. For [VOC], MAE and RMSE errors were 52.7 IAQI and 66.4 IAQI, respectively, demonstrating low error rates without significant deviations. Measurements for VOC ranged from 0 to 330 IAQI." [Gabriel and Auer, 2023]

"In conclusion, the [unseen transfer of the] LSTM model exhibits varying performance

Table 7.2: Evaluation metrics LSTM virtual sensing model transfer.

Pollutant	MAE	RMSE
CO_2	21.9	30.4
$PM_{2.5}$	0.3	0.6
VOC	52.7	66.4

in predicting indoor air pollutant concentrations for Office [4], with better results for CO_2 predictions and low error rates in terms of MAE and RMSE for [$PM_{2.5}$] and VOC predictions. However, there is room for improvement in capturing the variability of [$PM_{2.5}$] and VOC concentrations." [Gabriel and Auer, 2023]

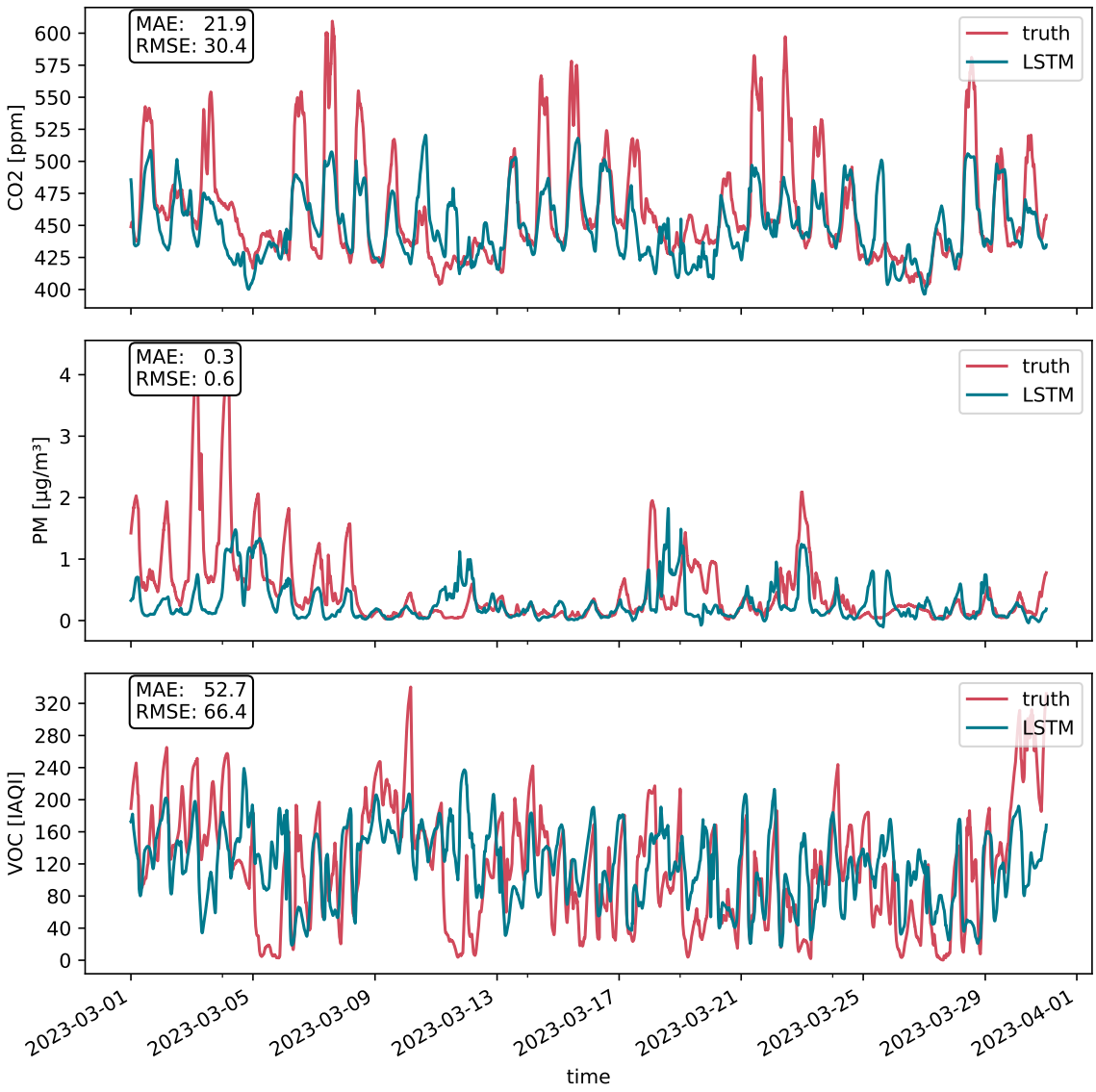


Figure 7.2: Comparison of Predicted (blue) and True (red) Time Series for Indoor Air Pollutants: VOC (bottom), $PM_{2.5}$ (middle), and CO_2 (top) [own representation]

7.3 LSTM transfer learning evaluation

In this section, the results of the transfer learning approach are evaluated. The section is divided into subsections for each room where the transfer learning was tested (Office 2, Office 3, Classroom). The section concludes with a summary and comparison of the results. To evaluate the predictive capability of the models, a comparison model (NULL-model) is introduced. The NULL-model is a dummy model that predicts the mean of the training data for each prediction. The NULL-model is used to compare the performance of the transfer learning models to a naive approach. If the model error metrics are lower than those of the NULL-model, it indicates some predictive capability of the model. Conversely, if the model error metrics are equal to or higher than those of the NULL-model, it indicates that the model is not able to make meaningful predictions.

7.3.1 Office 2

The figure displays the time series of prediction results for transferring the LSTM model to Office 2, separately for each pollutant. The plot shows measurements and predictions for October 2022. For CO_2, the LSTM model achieved a mean absolute error (MAE) of 30 ppm, while the NULL model obtained an MAE of 39 ppm. Thus, the LSTM model demonstrates a lower MAE than the NULL model, indicating its predictive capability for CO_2. The time series plots of the actual measurements and predictions for CO_2 show a correspondence in the trend and baseline concentration. However, the LSTM model exhibits a much lower variance and fails to accurately predict peak concentrations. For $PM_{2.5}$, the LSTM model achieved an MAE of 0.9 $\mu g/m^3$, whereas the NULL model obtained an MAE of 1.33. Therefore, the LSTM model performs better than the NULL model, indicating its predictive capability for $PM_{2.5}$. The time series plots show agreement in the concentration levels; however, the trend in $PM_{2.5}$ concentration is not always successfully predicted, and there are instances where pollutant peaks are shifted. An MAE of 40.4 was achieved for VOC, while the NULL model resulted in an MAE of 39.8. This indicates that the LSTM model performs comparably to the NULL model, suggesting a lack of predictive capability for VOC.

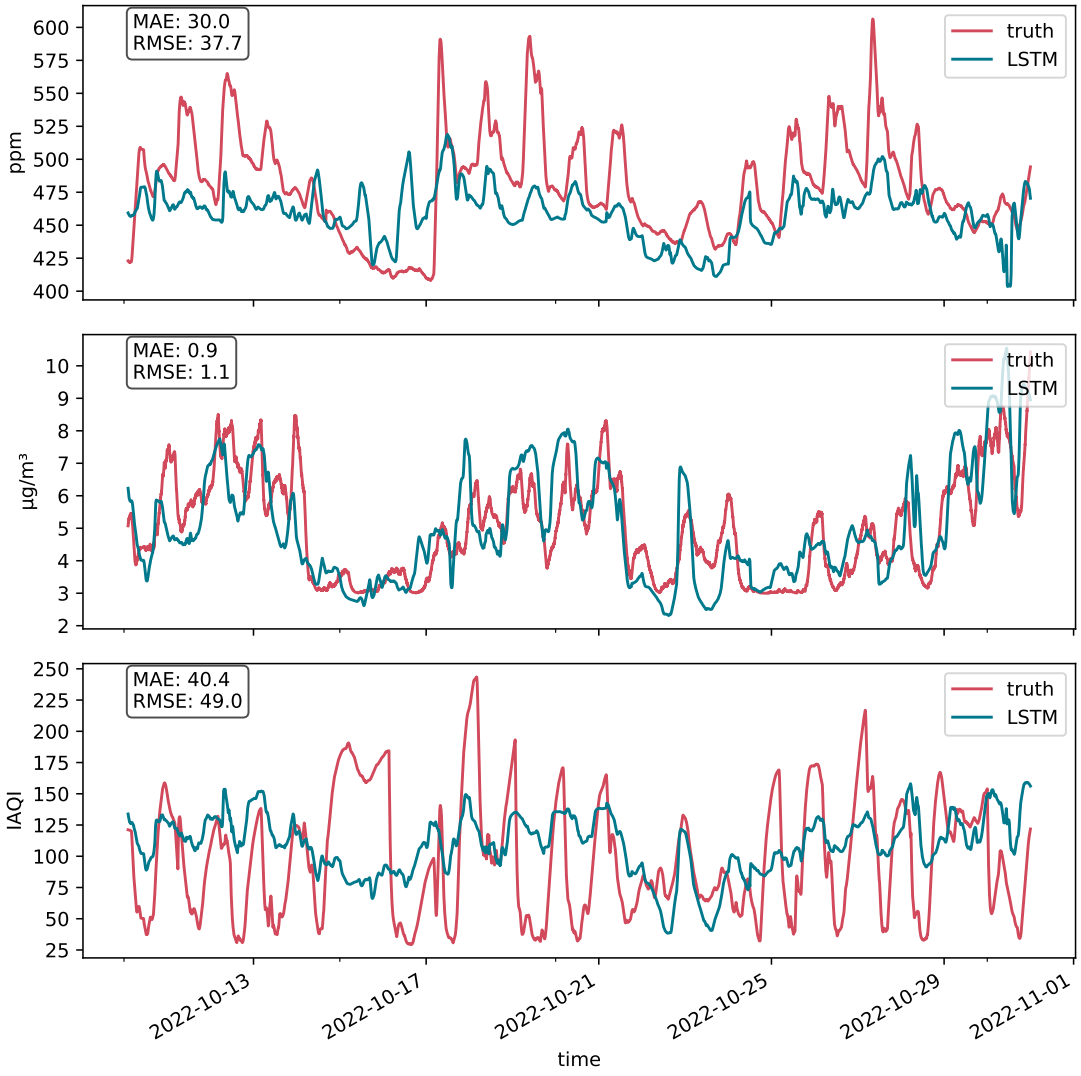


Figure 7.3: Comparison of LSTM model predictions and actual measurements for Office 2

7.3.2 Office 3

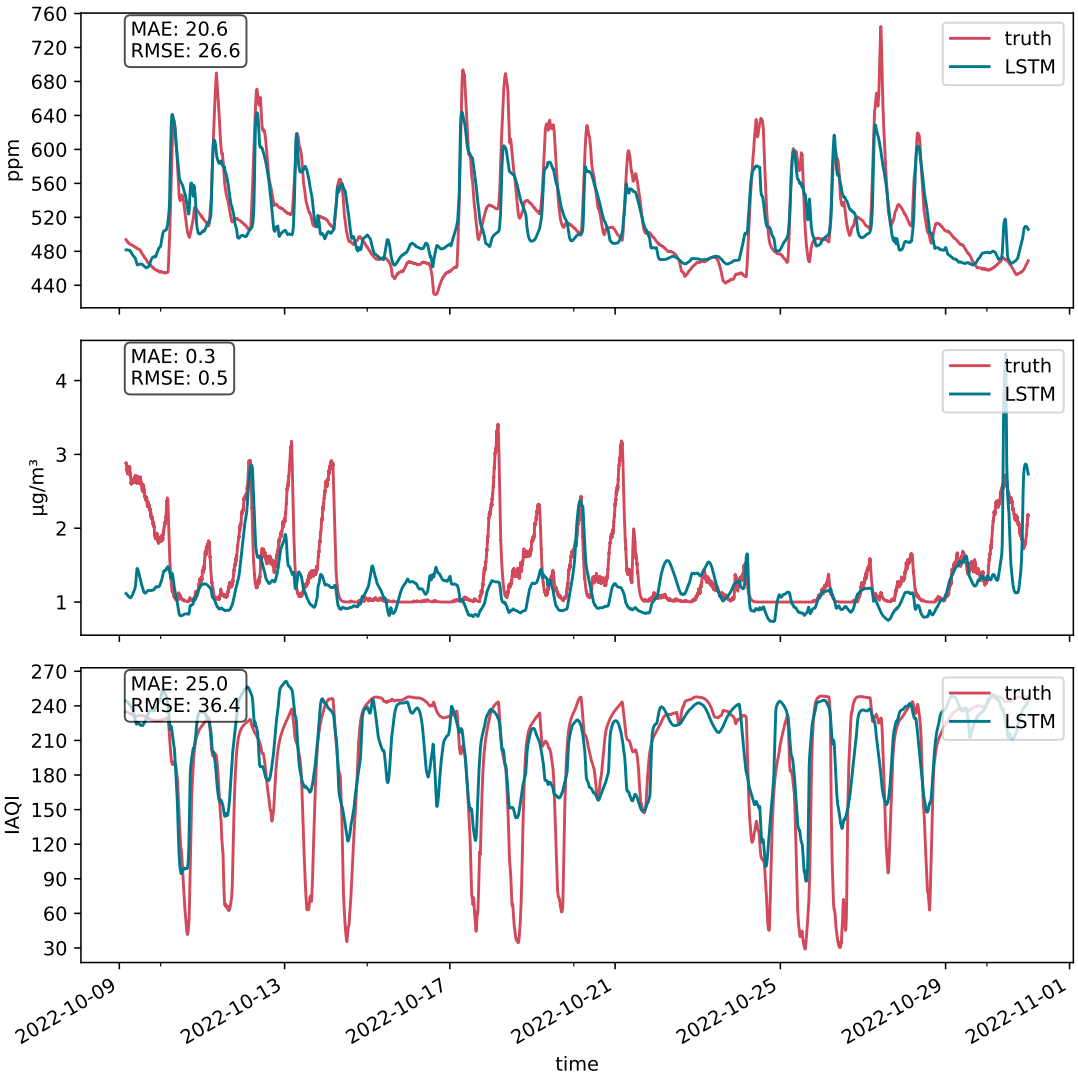


Figure 7.4: Comparison of LSTM model predictions and actual measurements for Office 3

The figure displays the time series of prediction results for transferring the LSTM model to Office 3. For CO_2, the LSTM model achieved an MAE of 20.6 ppm, while the NULL model obtained an MAE of 43 ppm. Therefore, the LSTM model significantly outperforms the NULL model, indicating its strong predictive capability for CO_2. The time series plots of the actual measurements and predictions for CO_2 show a strong correspondence, although the peaks in the measured values are slightly underestimated in the predictions. For $PM_{2.5}$, the LSTM model achieved an MAE of 0.3 $\mu g/m^3$, whereas the NULL model obtained an MAE of 0.5. Thus, the LSTM model performs better than

the NULL model, indicating its predictive capability for $PM_{2.5}$. The time series plots show agreement in the concentration levels. However, the trend in $PM_{2.5}$ concentration is not always successfully predicted, and there are instances where certain peaks are not recognized by the LSTM model. For VOC, an MAE of 25 was achieved by the LSTM model, while the NULL model resulted in an MAE of 57. This indicates that the LSTM model significantly outperforms the NULL model, indicating its strong predictive capability for VOC. The time series plot shows good agreement, with successfully detected peaks. However, the predictions tend to overestimate the concentration minima.

7.3.3 Classroom

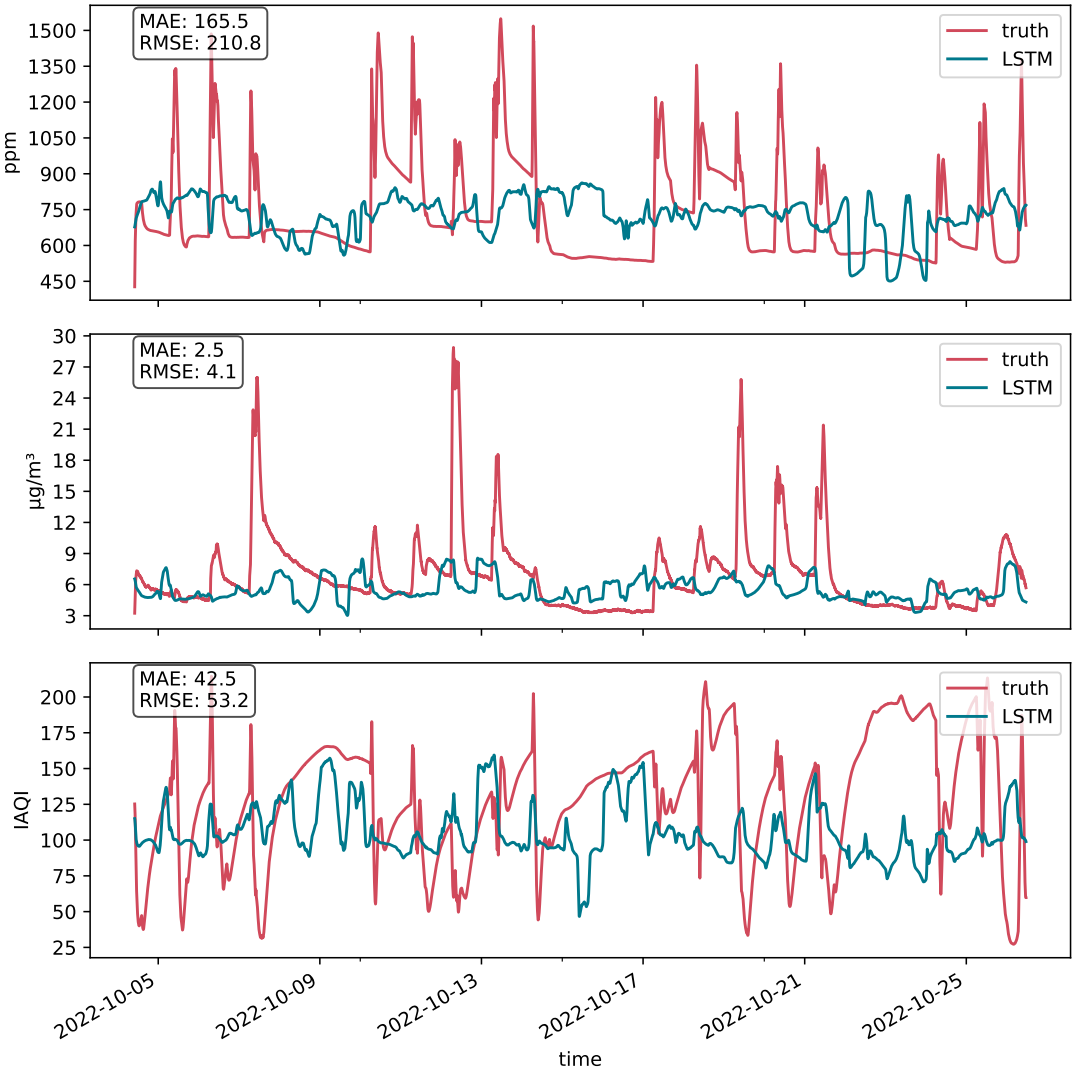


Figure 7.5: Time series prediction of LSTM model compared to ground truth

The figure displays the time series of the prediction results for transferring the LSTM model to the Classroom, separately for each pollutant. The plot presents the measurements and predictions for October 2022. For CO_2, the LSTM model achieved a mean absolute error (MAE) of 165.5 ppm, while the NULL model achieved a slightly higher MAE of 183 ppm. This indicates a weak predictive capability of the LSTM model for CO_2. The time series plots for CO_2 show no clear correspondence, suggesting that the LSTM model is not suitable for predicting CO_2. For $PM_{2.5}$, the LSTM model achieved an MAE of 2.5 $\mu g/m^3$, which is comparable to the MAE of the NULL model. This suggests that the LSTM model has no significant predictive capability for $PM_{2.5}$. For VOC, the LSTM model achieved an MAE of 42.5, while the NULL model had a slightly higher MAE of 46.2. This indicates a weak predictive capability of the LSTM model for VOC. The time series plots do not show a good agreement between the predicted and true values, indicating that the LSTM model cannot accurately predict VOC concentrations in the Classroom.

7.4 MLP model evaluation

To assess the performance of the LSTM model, it is compared with other machine learning algorithms. In this case, multi-layer perceptrons (MLP) and stochastic gradient descent (SGD) were selected for the comparison due to their coverage in the literature. The performance of the MLP and SGD models was evaluated in Office 1, Office 2, and Office 3, as well as the Classroom for both the training room and the transfer learning approach.

7.4.1 Office 1

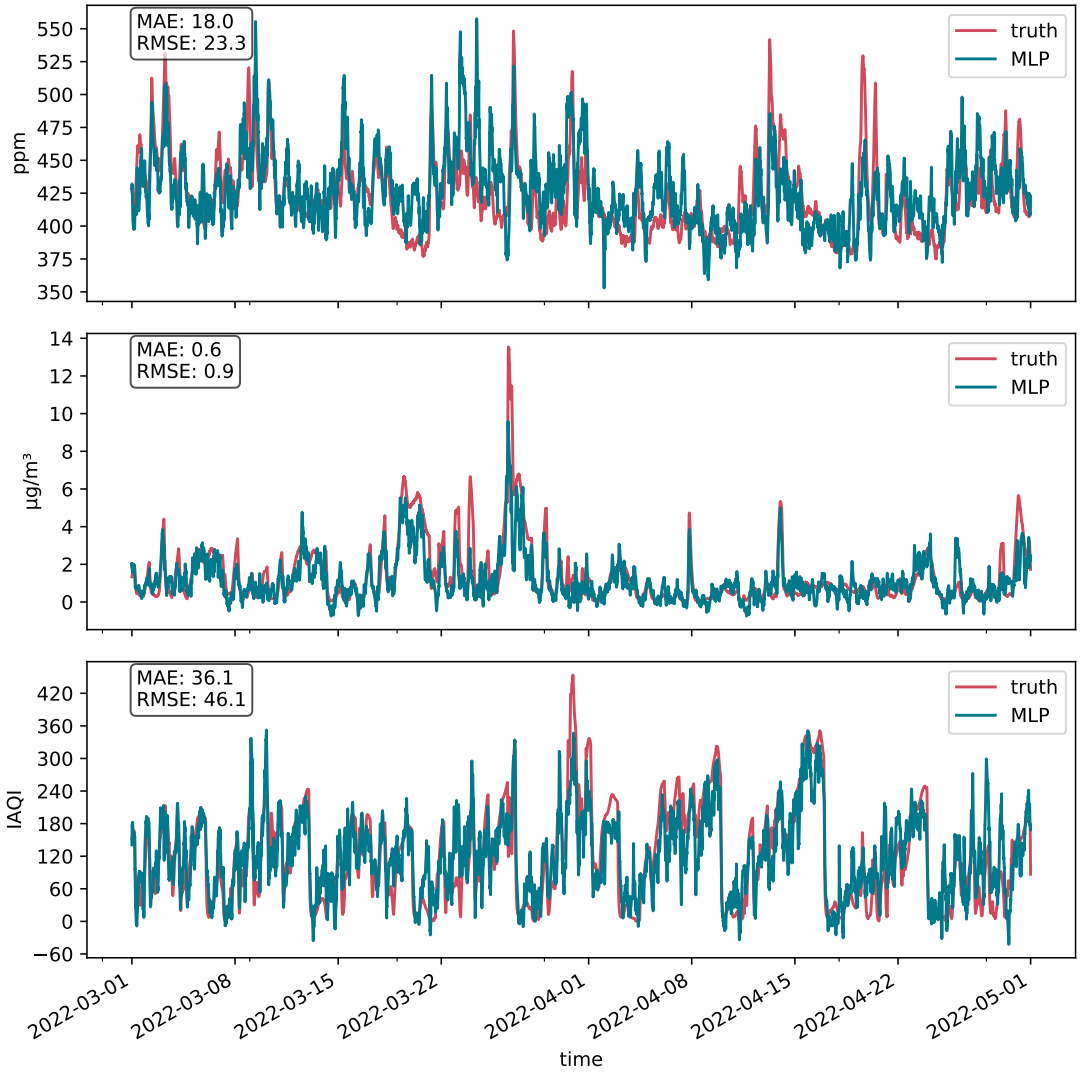


Figure 7.6: Office 1: Time series prediction of MLP model compared to ground truth

The figure illustrates the time series of the MLP model for Office 1.An MAE of 18 ppm was achieved for CO_2, which is lower than the MAE of 24 ppm obtained by the NULL model. This lower MAE indicates that the MLP model has a predictive capability for CO_2. The time series plots of the measured values and the predictions for CO_2 show good correspondence. However, the predictions exhibit higher volatility compared to the measured values, with some peak concentrations being overestimated and others being underestimated. For $PM_{2.5}$, the MLP model achieved an MAE of 0.6 $\mu g/m^3$, while the NULL model achieved 0.99. This indicates that the MLP model is better than the NULL model in terms of predictive capability for $PM_{2.5}$. The time series plots also demonstrate a high level of accordance between the measured values and the predictions. However, the predictions show a higher volatility of concentrations compared to the measured values. In the case of VOC, the LSTM model achieved an MAE of 36.1, while the NULL model resulted in an MAE of 72.5. This significant improvement indicates that the LSTM model has a predictive capability for VOC. The time series plots for VOC show good correspondence between the measured values and the predictions. However, the predictions exhibit higher volatility than the measured values, resulting in some peak concentrations being overestimated.

7.4.2 Office 2

The figure displays the time series for the prediction results of transferring the MLP model to Office 2 for each pollutant separately.
For CO_2, an MAE of 30.6 ppm was achieved, whereas the NULL model had an MAE of 39 ppm. Therefore, the MLP model outperforms the NULL model, indicating its predictive capability for CO_2. The time series plots of the ground truth and the MLP prediction for CO_2 exhibit correspondence in terms of trend and baseline concentration. However, the MLP predictions show high volatility, resulting in significant underestimations.
For $PM_{2.5}$, an MAE of 1.4 $\mu g/m^3$ was achieved. This performance is comparable to the NULL model, suggesting that the MLP model lacks predictive capability for $PM_{2.5}$.
An MAE of 42.1 was achieved for VOC, while the NULL model yielded an MAE of 39.8. Thus, the MLP model performs worse than the NULL model, indicating its lack of predictive capability for VOC.

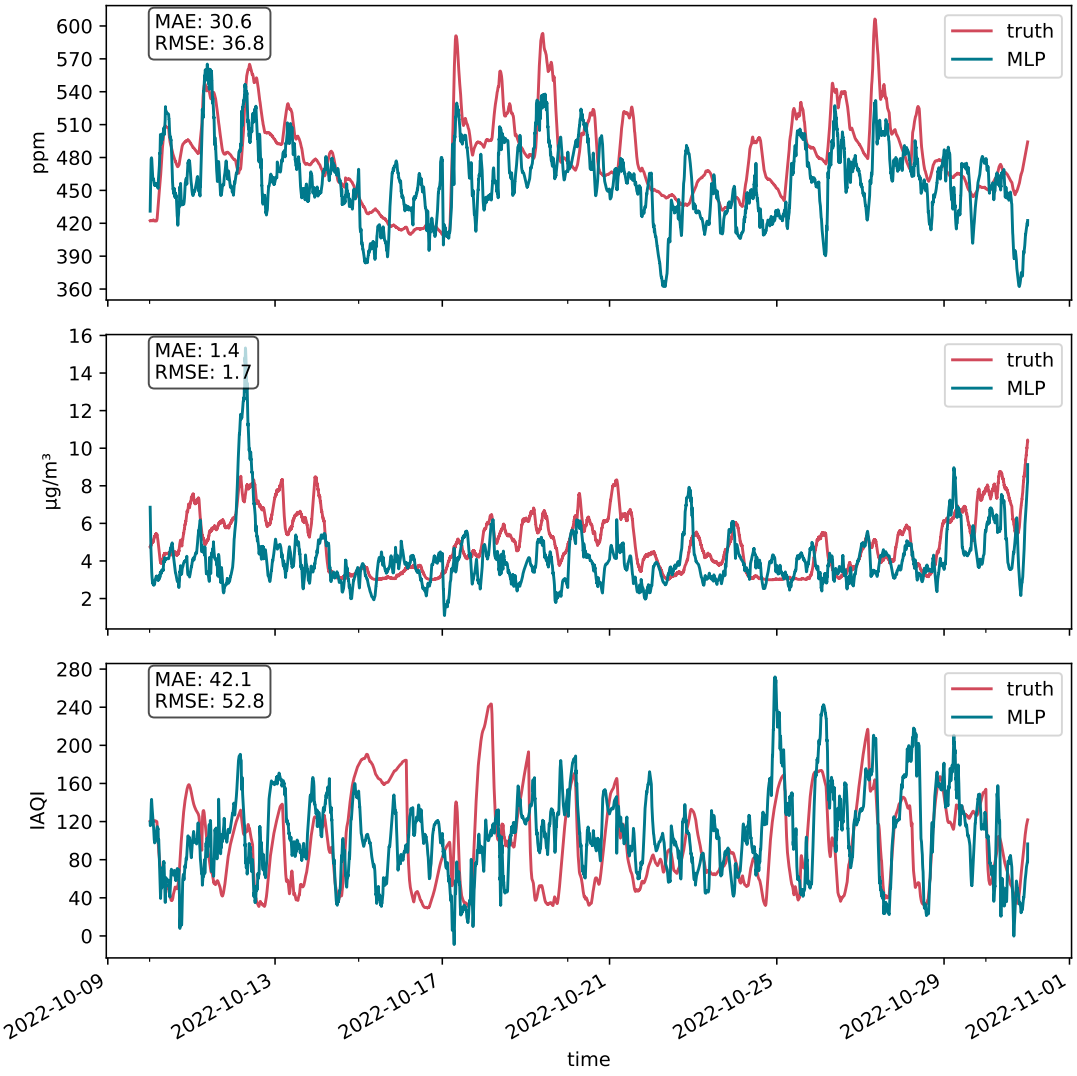


Figure 7.7: Office 2: Time series prediction of MLP model compared to ground truth

7.4.3 Office 3

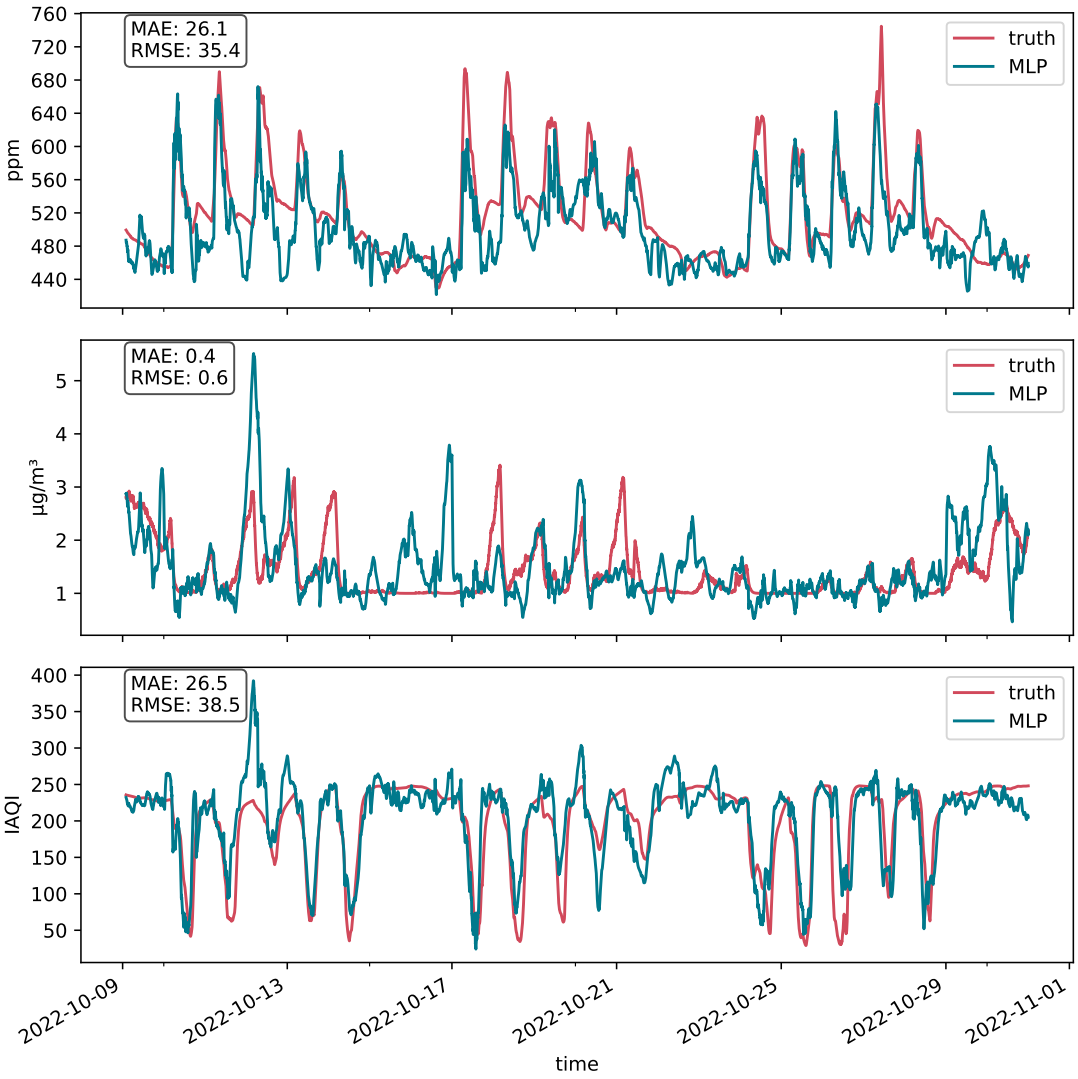


Figure 7.8: Office 3: Time series prediction of MLP model compared to ground truth

The figure illustrates the time series for the prediction results of transferring the MLP model to Office 3.
For CO_2, an MAE of 26.1 ppm was achieved, whereas the NULL model had an MAE of 43 ppm. Thus, the MLP model outperforms the NULL model, indicating its predictive capability for CO_2. The time series plots of the ground truth and the MLP prediction for CO_2 exhibit strong correspondence. However, volatility in the MLP model's predictions introduces errors.
For $PM_{2.5}$, an MAE of 0.4 $\mu g/m^3$ was achieved, while the NULL model achieved 0.5. The

MLP model performs better than the NULL model, indicating its predictive capability for $PM_{2.5}$. The time series plots demonstrate agreement between the prediction results and measurements during certain periods. However, the model incorrectly identifies concentration peaks during periods of low pollution and fails to identify peaks during periods of high pollution.
An MAE of 26.5 was achieved for VOC, whereas the NULL model resulted in an MAE of 57. Therefore, the MLP model significantly outperforms the NULL model, indicating its strong predictive capability for VOC. The time series plot displays good agreement, with peaks being detected, even though they are regularly overestimated.

7.4.4 Classroom

The graph displays the time series prediction results of the MLP model for the Classroom, focusing on each pollutant individually. For CO_2, the model achieved a mean absolute error (MAE) of 176.5 ppm, while the NULL model had a slightly higher MAE of 183 ppm. This suggests that the MLP model has limited predictive capability for CO_2. The time series plots of the ground truth and the model prediction for CO_2 show that the model was able to identify some pollutant peaks. However, the overall error is large, especially during lower concentrations. For $PM_{2.5}$, the MAE was calculated to be 3.2 $\mu g/m^3$, while the NULL model obtained an MAE of 2.5. Thus, the LSTM model performs worse than the NULL model, indicating that it has no predictive capability for $PM_{2.5}$. For VOC, an MAE of 53.5 was achieved, while the NULL model resulted in an MAE of 46.2. This indicates that the MLP model also has no predictive capability for VOC pollution.

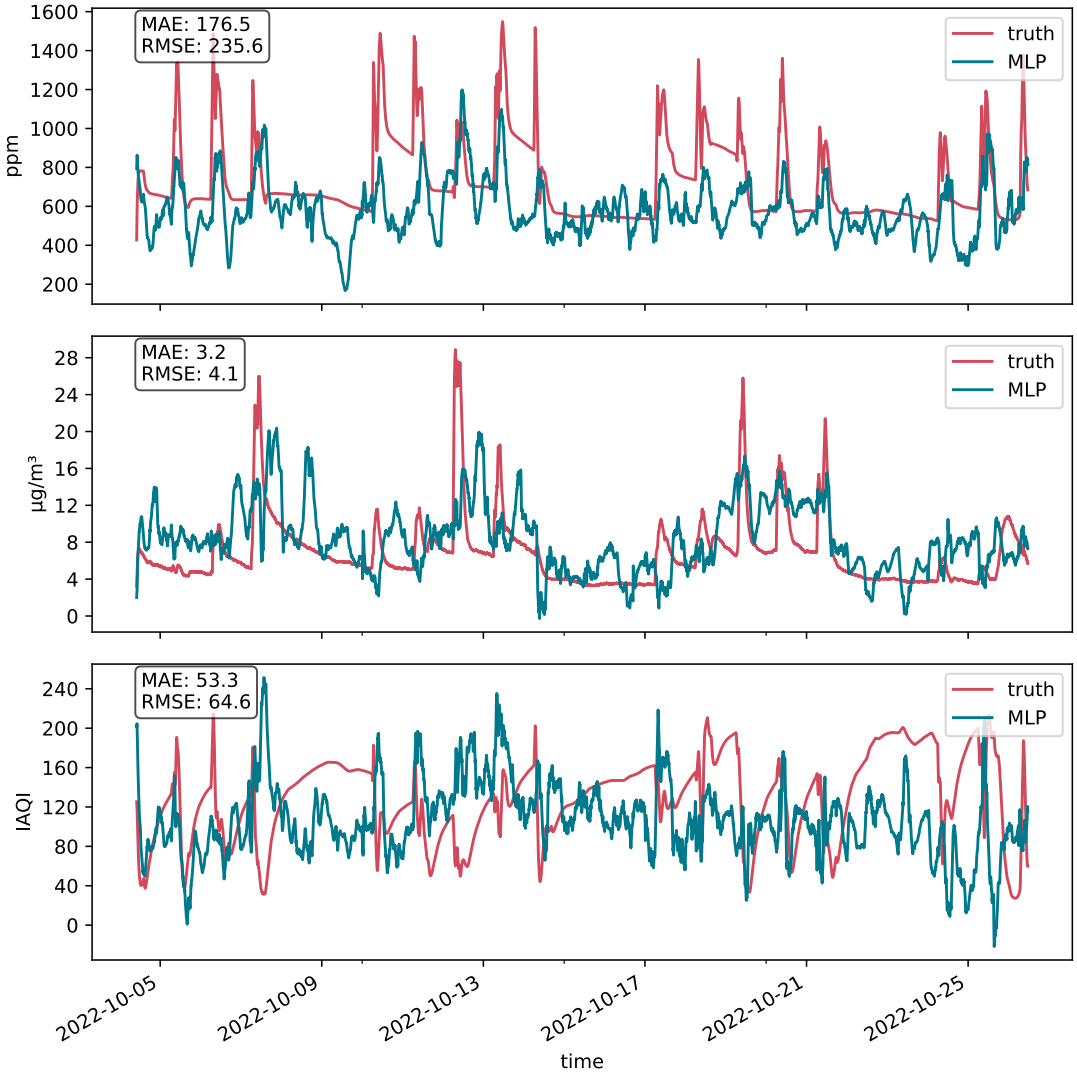


Figure 7.9: Classroom: Timeseries prediction of MLP model compared to ground truth

7.5 SGD model evaluation

7.5.1 Office 1

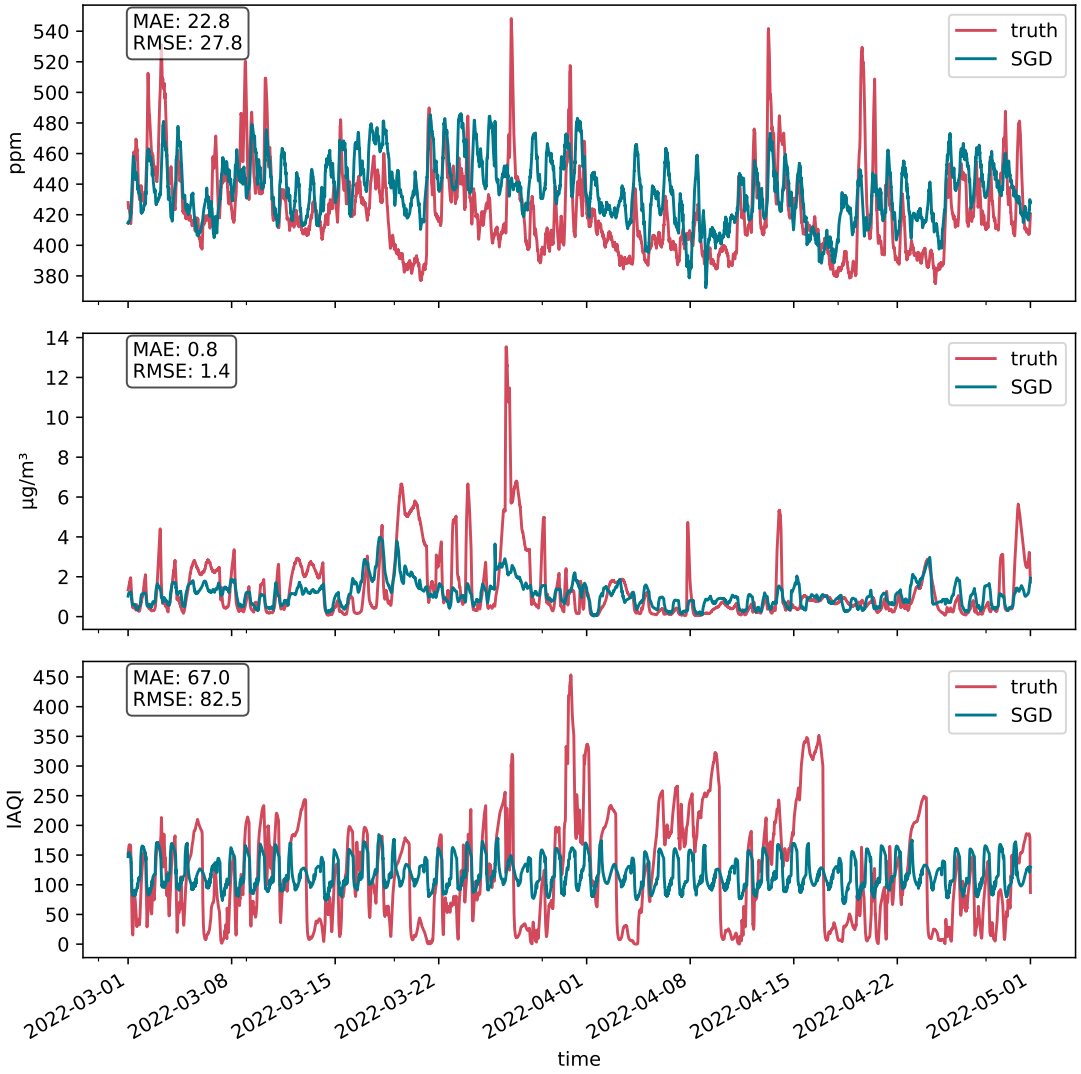


Figure 7.10: Office 1: Timeseries prediction of SGD model compared to ground truth

The graph above illustrates the time series plots for the SGD model. For CO_2, the model achieved an MAE of 22.8 ppm, while the NULL model had a slightly worse MAE of 24 ppm. Therefore, the predictive capability of the SGD model for CO_2 is weak. The time series plots of the ground truth and the model prediction for CO_2 show correspondence in terms of concentration increase or decrease. However, there is often an offset between the model and the truth, leading to significant overestimations

during certain periods. Additionally, the variance of the prediction results was lower than that of the measured values, resulting in underestimations of peak concentrations and overestimations of minima. For $PM_{2.5}$, the model achieved an MAE of 0.8 $\mu g/m^3$, while the NULL model obtained 0.99. The SGD model is only slightly better than the NULL model, indicating weak predictive capability for $PM_{2.5}$. The time series plots show high accordance during periods of low concentration. However, the SGD model fails to accurately predict high pollutant events and significantly underestimates concentrations during these events. For VOC, an MAE of 67 was achieved, while the NULL model resulted in an MAE of 72.5. Thus, the SGD model performs only slightly better than the NULL model, indicating weak predictive capability for VOC. However, the time series plots of the ground truth and the model prediction for VOC show correspondence in terms of the increase and decrease of VOC concentrations, albeit with lower variance in the SGD predictions. Therefore, peaks were underestimated and minima were overestimated.

7.5.2 Office 2

The figure displays the time series prediction results for transferring the SGD model to Office 2, separately for each pollutant. For CO_2, an MAE of 25.4 ppm was achieved, compared to the NULL model's MAE of 39 ppm. The lower MAE of the SGD model in comparison to the NULL model indicates its predictive capability towards CO_2. The time series plots for CO_2 show good agreement between SGD predictions and measured truth, although peak concentrations are regularly underestimated. For $PM_{2.5}$, an MAE of 1.0 $\mu g/m^3$ was achieved, while the NULL model achieved 1.33. Therefore, the SGD model performs better than the NULL model, indicating its predictive capability towards $PM_{2.5}$. The time series plots show general accordance in the concentration baseline, but spikes in $PM_{2.5}$ could not be reliably predicted.
For VOC, a MAE of 37.4 was achieved, compared to the NULL model's MAE of 39.8. The LSTM model only slightly outperforms the NULL model, indicating a weak predictive capability towards VOC. The time series plots show that the model successfully generalizes pollutant increase and decrease, but the prediction has a significantly lower variance and is offset.

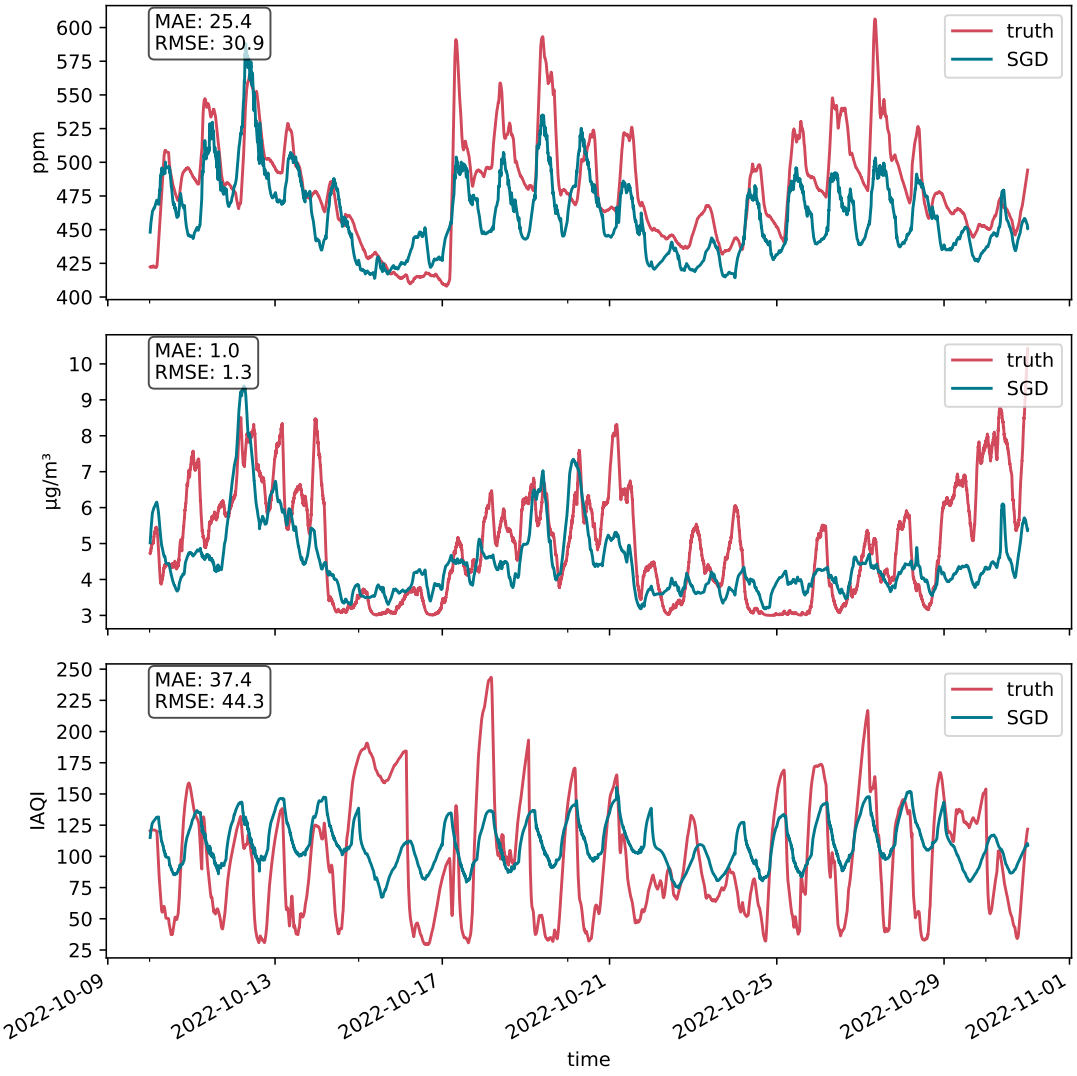


Figure 7.11: Office 2: Time series prediction of SGD model compared to truth

7.5.3 Office 3

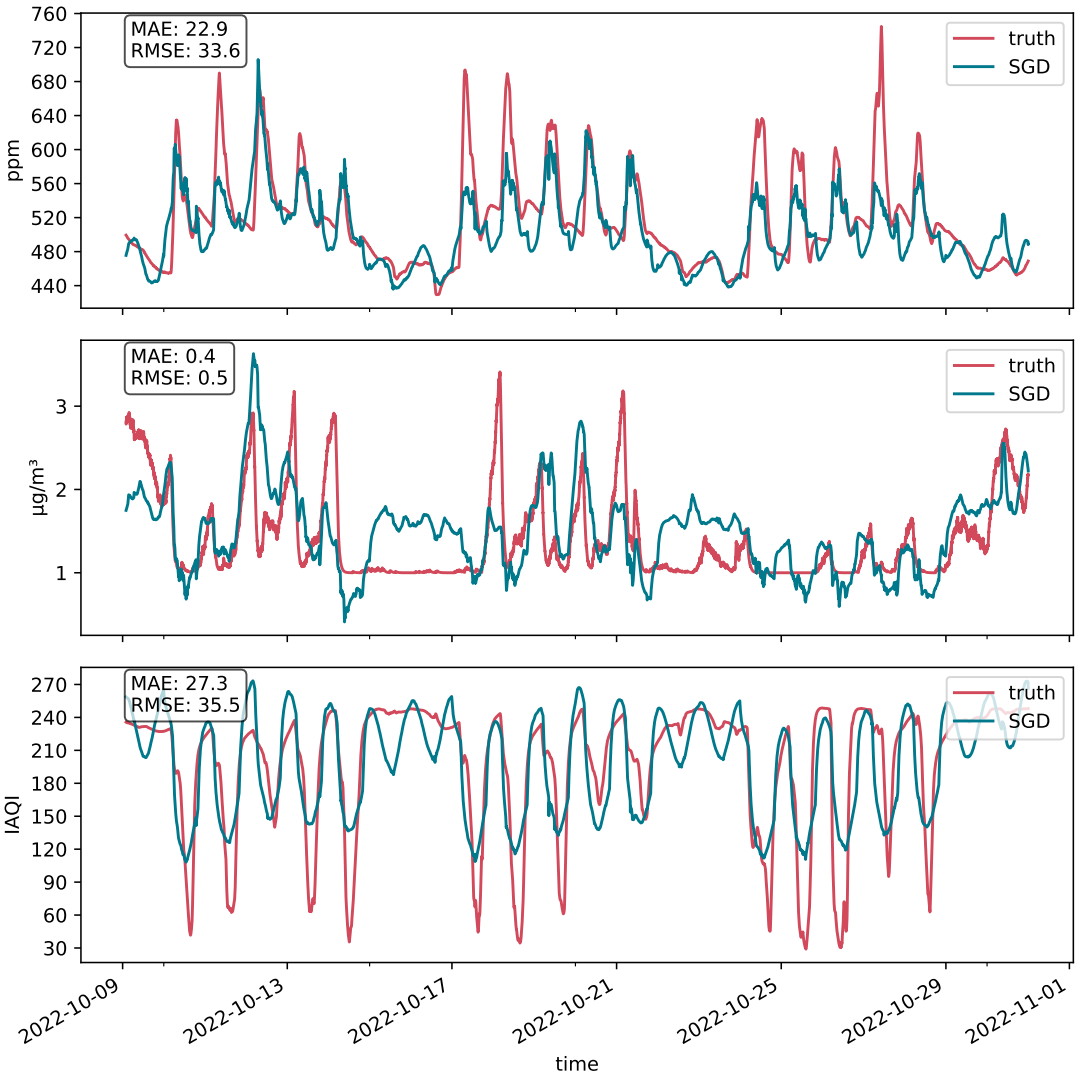


Figure 7.12: Office 3: Time series prediction of SGD model compared to truth

The figure displays the time series prediction results for the transfer of the SGD model to Office 3.
For CO_2, an MAE of 22.9 ppm was achieved, compared to the NULL model's MAE of 43 ppm. The SGD model performs significantly better than the NULL model, indicating its strong predictive capability towards CO_2. The time series plots for CO_2 show strong correspondence, although peaks in measured values are slightly underestimated in the predictions.
For $PM_{2.5}$, an MAE of 0.4 $\mu g/m^3$ was achieved, while the NULL model achieved 0.5.

Therefore, the SGD model outperforms the NULL model, indicating its predictive capability towards $PM_{2.5}$. The time series plots show that the SGD model's predictions cannot reliably detect pollution peaks and partially overestimate concentration during low pollution periods.
An MAE of 27.3 was achieved for VOC, compared to the NULL model's MAE of 57. The SGD model performs significantly better than the NULL model, indicating its strong predictive capability towards VOC. The time series plot shows good agreement, with peaks successfully detected. However, the prediction overestimates the concentration minima.

7.5.4 Classroom

The figure depicts the time series for the prediction results of transferring the SGD model to the Classroom for each pollutant individually.
For CO_2, the SGD model achieved a mean absolute error (MAE) of 125.4 ppm, while the NULL model achieved an MAE of 183 ppm. This suggests that the SGD model has predictive capability for CO_2. The time series plots for CO_2 show good agreement between the predicted and actual concentration baseline, and the model is able to capture increases and decreases in concentration. However, the SGD predictions significantly underestimate peak concentrations.
For $PM_{2.5}$, the SGD model achieved an MAE of 1.6 $\mu g/m^3$, while the NULL model achieved an MAE of 2.5. This indicates that the SGD model has predictive capability for $PM_{2.5}$. The time series plot shows agreement between the predicted and measured values during periods of low pollutant levels, but the SGD model underestimates pollutant peaks.
For VOC, the SGD model achieved an MAE of 36.8, while the NULL model achieved an MAE of 46.2. This suggests that the SGD model has predictive capability for VOC. The time series plots demonstrate that the model can predict changes in VOC concentration. However, there is reduced variance in some time periods, leading to measurement errors due to underestimation and overestimation of VOC concentration.

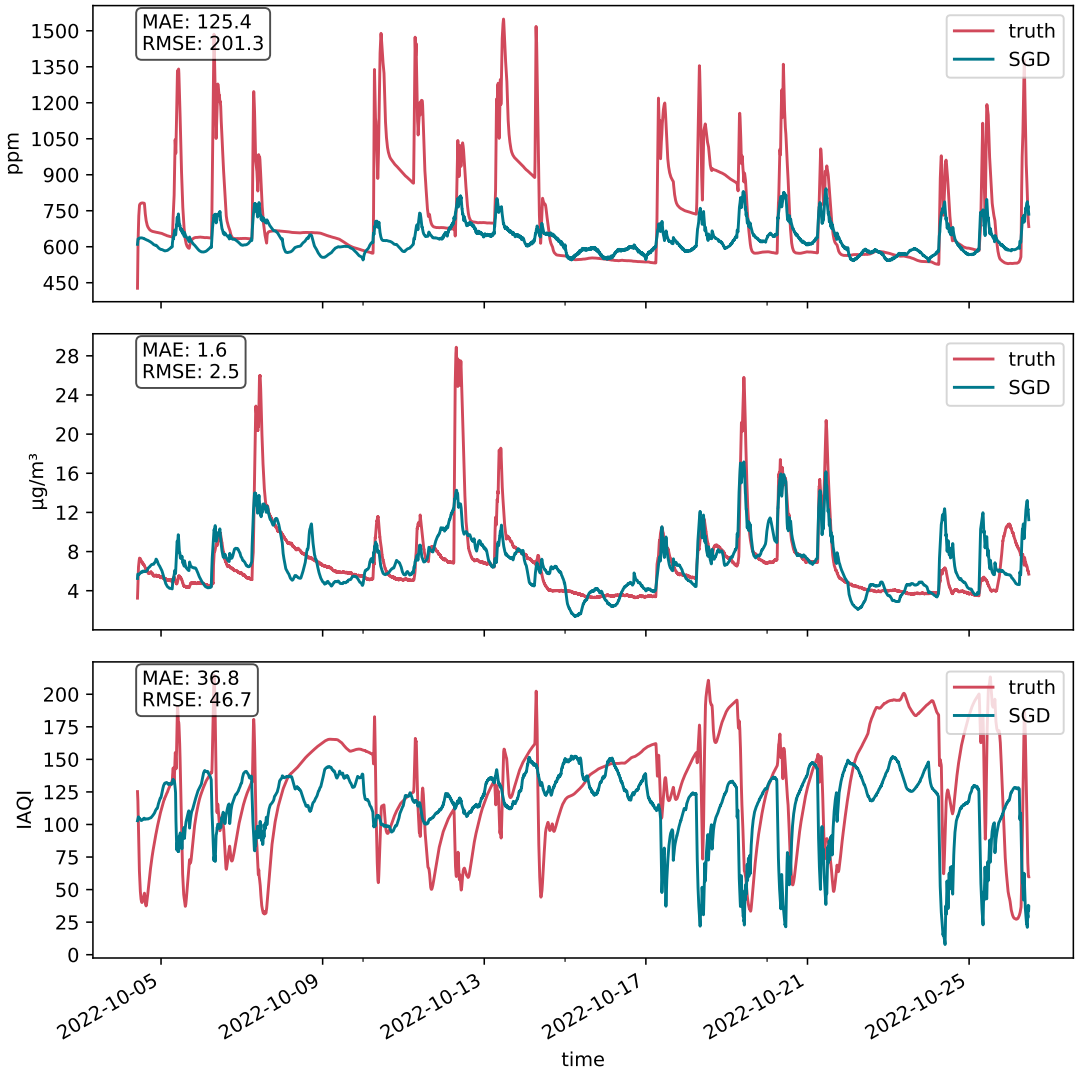


Figure 7.13: Classroom: Timeseries prediction of SGD model compared to truth

7.6 Model comparison

In the following the performance of the machine learning models (LSTM, MLP, SGD) in predicting indoor air pollutant concentrations, specifically CO_2, $PM_{2.5}$, and VOC is compared. These models were trained in Office 1 and then applied to Office 2, 3, 4 and the Classroom using a transfer learning approach. The Mean Absolute Error (MAE) results, as shown in Table 7.3 summarizes the evaluation of the models.
In the context of Office 1, the LSTM model performs best for all pollutants, achieving the lowest MAE values for CO_2, $PM_{2.5}$, and VOC. LSTM significantly outperforms the MLP and SGD models, underscoring the efficacy of LSTM for time series application were a large volume of training data is available.
In the testing rooms all models exhibit lower performance compared to the training room. However, the trained and tuned LSTM model still performs significantly better than the NULL model in the office rooms indicating its transferability. The MLP model performs less effectively when applied to other office rooms and only slightly surpasses the NULL model. Notably, the linear model (SGD) also demonstrates comparable or slightly superior performance to the more complex LSTM and MLP models in the testing rooms, which is attributable to the simplicity of the model and its ability to generalize effectively with limited training data. Prediction results are significantly diminish for the Classroom for all models. This discrepancy is attributed to the differences in building characteristics, operation, and occupancy patterns between the Classroom and the Office rooms. The dynamics and occupancy density of the Classroom couldn't be adequately generalized using training data from lower occupancy office environments, resulting in higher error rates for the Classroom setting.
Furthermore, unseen transfer without applying the transfer-learning approach was tested for the LSTM model, by transferring the model to the previously unseen Office 4, with MAE results of 21.9 (CO_2), 0.3 ($PM_{2.5}$), and 52.7 (VOC) notably achieving lower error rates for CO_2 and $PM_{2.5}$ compared to the transfer learning approach for the other Office rooms. This indicates that the low volume of tuning data applied to the LSTM during the transfer learning approach led to detrimental effects on the model performance possibly due to overfitting to a limited data set.
In conclusion, the LSTM model surpasses MLP and SGD if sufficient training data is available. However, the SGD model is a viable alternative if training data is limited.

Table 7.3: Comparing the MAE results of different ML algorithms - best results marked in bold

Office	Parameter	NULL	LSTM	MLP	SGD
Office 1	CO_2	26.09	**15.4**	18.64	22.48
	$PM_{2.5}$	0.71	**0.3**	0.5	0.59
	VOC	60.9	**20.1**	37.26	58.28
Office 2	CO_2	38.65	32.98	30.05	**25.97**
	$PM_{2.5}$	1.52	**0.92**	1.38	0.95
	VOC	50.21	47.11	46.33	**44.59**
Office 3	CO_2	41.14	29.41	29.27	**25.98**
	$PM_{2.5}$	0.66	**0.53**	0.54	0.53
	VOC	53.99	30.41	**28.93**	30.7
Classroom	CO_2	495.48	**408.94**	481.09	433.58
	$PM_{2.5}$	2.76	2.49	3.88	**2.37**
	VOC	49.86	**44.75**	54.92	45.25

7.7 Applicability of ML based virtual indoor air pollutant sensors

"[The results] demonstrate the potential of machine learning models, specifically [Long Short-Term Memory] networks , to accurately predict indoor air pollutant concentrations in a range of environments. Using a large dataset with several years of accumulated data, a virtual indoor air pollutant sensor [was built] that exhibited strong performance in predicting indoor air pollutant concentrations for the room in which it was trained. The evaluation results indicated a very high correlation between the actual and predicted pollutant concentrations for [$PM_{2.5}$] and VOC, with performance metrics MAE = 0.3 $\mu g/m^3$ and RMSE = 0.5 $\mu g/m^3$ [...] for [$PM_{2.5}$]; and MAE = 20.1 IAQI [and] RMSE = 31.4 IAQI [...] for VOC. The [visual analysis] show[ed] that the model [could] identify most pollutant peaks during the testing period with only a slight underestimation of peak concentrations. For CO_2, the model achieved less ideal but reasonable prediction results. The performance metrics of MAE = 15.4 ppm [and] RMSE = 20.2 ppm for CO_2 indicated very low errors with insignificant outliers [...]. [However, visual analysis showed, that the model could] not explain the variability of the actual concentrations and showed some erroneous predictions, such as misidentified pollutant peaks during the testing period. Nevertheless, the predictions resulted in a mean absolute error within the range of the measurement inaccuracy of most sensors. When [testing the unseen] transfer [of] the model to [Office 4], the LSTM model demonstrated varying performance, with better results for CO_2 [...]. [$PM_{2.5}$] and VOC predictions [achieved low error rates]. However, there is room for improvement in capturing the variability of [$PM_{2.5}$] and VOC concentrations." [Gabriel and Auer, 2023] Despite these challenges, the LSTM model shows its potential in generalizing its ability to predict indoor air pollutant concentrations

in different rooms.
Furthermore, a transfer learning approach was tested that used a short tuning period to adapt the model to the changed environment in other rooms, buildings, and typologies. While transferring the pre-trained models to other rooms resulted in lower performance compared to the training rooms, acceptable errors were achieved in rooms of the same typology. However, when transferring the model to a different typology, such as a Classroom, model performance significantly decreased, and the models achieved no meaningful predictions.
In comparing LSTM, SGD, and MLP models applied to varying testing environments, the LSTM model outperformed the MLP model in its generalization ability. LSTM and MLP models proved effective within the same room they were trained in, but LSTM demonstrated superior adaptability when transferred to other rooms. However, these complex models exhibited high error rates when predicting indoor air pollutant concentrations in classrooms, suggesting struggles with environments of different occupancy patterns and dynamics.
Contrastingly, despite its simplicity, the SGD linear model performed at a similar level to the more complex models in environments with limited training data, such as the transfer learning for Office 2 and Office 3, even though the SGD model performed significantly worse than the LSTM and the MLP model in the training room. This indicates that while LSTM performed significantly better in Office 1, where high amounts of training data were available, simple models like SGD may sometimes match or exceed the performance of more complex models if only a little or error-prone data are available.
The model evaluation indicates that machine learning models, particularly Long Short-Term Memory (LSTM) networks, are effective in predicting indoor air pollutants, as demonstrated by the low error rates achieved in the testing set of the room. The unseen transfer and transfer-learning evaluation showed diminished performance of the model in new rooms, due to the limited training dataset, which consists solely of data from one room. Consequently, it is crucial to enhance the unseen transfer and performance of the virtual sensing models by generating larger and more diverse datasets. By training on more data, the generalizability of this model is expected to improve further.

Part IV

Case studies

8 Case Study 1 - Open Office

The case studies examine the application of the virtual sensing model in demand-controlled ventilation in two different environments: an open office and a classroom. Each case study provides a detailed description of the setup, framework, calibration, control strategies, and results. In case study 1, a digital twin of Office 1 (see Section 4.2.1) is generated using transient thermal simulations calibrated with measurements from the aforementioned dataset. The following chapter provides a comprehensive overview of the methods and results of case study 1, specifically focusing on the open office scenario.

8.1 Simulation Framework

This section describes the creation of the model, which is built and simulated using TRNSYS. To ensure automation, the inputs are parametrized. The parametrization for this study is implemented in the TRNSYS Grasshopper interface, TRNLizard. TRNLizard is a Python-based interface between TRNSYS and the graphic programming environment, Grasshopper. It allows for a modular approach and increased expandability. TRNLizard enables the transfer of geometries and parameters to the TRNSYS simulation engine by creating simulation files based on predefined templates. The simulations for the pre-written simulation files are launched via command-line arguments. Although TRNLizard currently offers less functionality compared to TRNSYS Studio, its open-source implementation allows experienced users to develop new tools and functions. To integrate the virtual sensing model, a co-simulation framework is required to exchange indoor environment data between the models at every timestep. In this co-simulation, the building energy simulation is combined with the virtual indoor air pollutants sensor. The virtual indoor air pollutant sensor, implemented in Python, is called by the TRNSYS engine at every simulation step. It returns the current pollutant values to TRNSYS, which are then integrated into the control algorithms to determine the supply air volumes. The co-simulation is based on the TRNSYS Type 3157 [Jacob, 2012]. Figure 8.1 provides a schematic representation of this workflow.

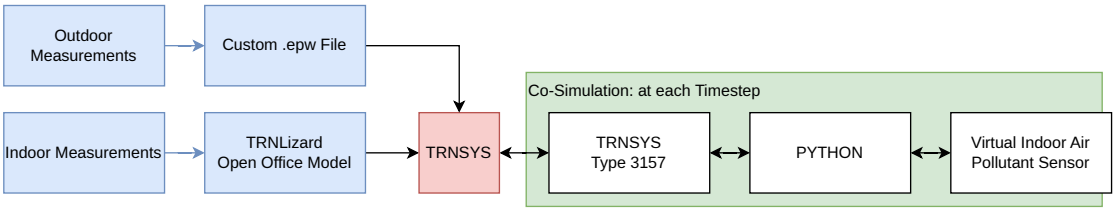


Figure 8.1: Conceptual simulation framework

8.2 Simulation Setup

This section outlines the setup of the simulation model, including a description of the modeled zone, its environment, building technology, and building materials.

8.2.1 Location and Model

The thermal simulation is conducted for Office 1, a single open office located in a high-rise building. The building was constructed in 2011 and has a total floor area of 111,000 square meters. The open office occupies an area of 394 square meters. The maximum occupancy for the office is 35; however, due to the COVID-19 pandemic and part-time home office arrangements, the maximum occupancy throughout the observational period was reduced to 18. The office has two external facades oriented towards the southeast and northwest. The other boundaries of the room are adiabatic, as they are connected to other office spaces.

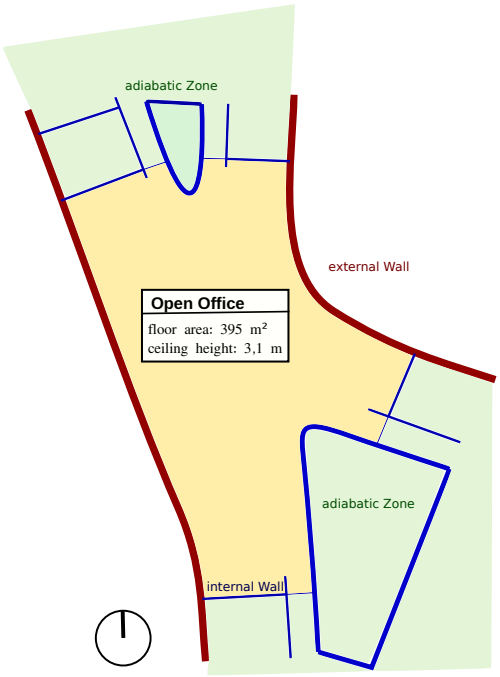


Figure 8.2: Schematic simulation setting

8.2.2 Building technology

The building technology of the examined building is summarized in detail by Auer et al. [2020]. The key facts are as follows. The heating system is supplied by geothermal heat pumps and connected to the district heating grid. The cooling system of the building is supplied by geothermal groundwater heat pumps, which are supported by additional

cooling compressors. Fresh air is supplied by a central mechanical ventilation unit with regenerative and recuperative heat recovery. Additionally, natural ventilation via operable windows is available in the lower floors. The mechanical ventilation operates constantly between 5:15 am and 8 pm at an air change rate of 1.6 air changes per hour, resulting in a supply air volume of 1933 m^3/h for Office 1. Fresh air is introduced through ground-level induction units and extracted at the ceiling level. The HVAC system is not designed to provide heating and cooling, but rather to supply air at an indoor air temperature within the range of 20 °C to 22 °C. Heating and cooling are supplied to the room through a concrete core activation system and radiators for peak heating demand in winter. The operating mode of the concrete core activation system is determined based on the 24-hour average outdoor temperature. Cooling is activated when the average outdoor temperature exceeds 19°C, while heating is activated when the average outdoor temperature falls below 15°C. The concrete core activation operates in heating mode between 6 am and 10 pm. To reduce daytime peak loads, the concrete core activation is also operated in cooling mode and during night hours. The activation and deactivation of the concrete core activation system are controlled by a surface temperature setpoint. The room is equipped with a movable outside shading system, which is automatically controlled based on global radiation but can be overridden by the user.

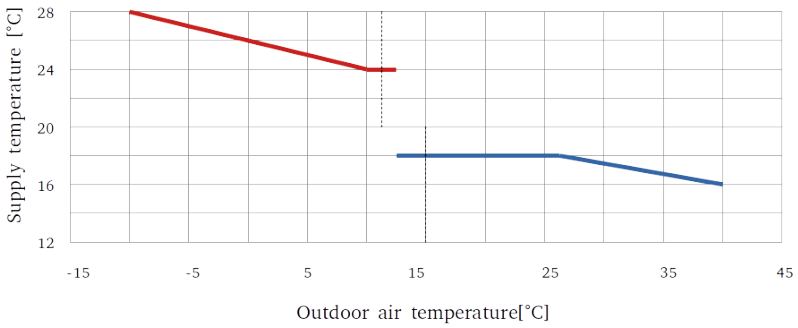


Figure 8.3: Exemplary TABS heating and cooling curve (own representation; based on [Hepf et al., 2023])

8.2.3 Building materials

The building is constructed of concrete, and the outer walls are insulated. The internal walls, external walls, and ceilings are made of heavy construction materials, exposing thermal mass to the room's interior. Due to the concrete core activation system, the ceiling is exposed, and air supply and exhaust openings are located in the interior walls rather than the ceilings. The two external facades have a concrete parapet up to 0.9 m, followed by a ceiling-height window front. The insulation level of the building was determined according to the then-valid building code EnEV (Energieeinsparverordnung). The requirements of the building code specified a U-Value of 0.28 W/m^2K for the external

walls and a U-Value of 1.3 W/m²K for the external windows. The building parts in the simulation were modeled based on these values, with a simple two-layer construction consisting of concrete and external insulation for the external building parts.

8.2.4 Model calibration using measured values

To incorporate the measurements into the calibrated simulation model, custom schedules were created: Air Change Rate, Electric lighting, Window opening, Occupancy (binary), and Occupant count. These variables were integrated by resampling the input signals derived from measurement data to match the time resolution of the simulation (15 minutes). This ensures that the model accurately reflects real-world conditions. The electric lighting and occupancy count were generated by combining measurement values. For electric lighting, the input data for Illumination and Outdoor Radiation were combined to compute the electric lighting component. The occupancy count was derived from the CO_2 concentration and the Occupancy signal.

8.3 Customized weather file

To facilitate calibration for the observational period, a customized weather file (epw) was generated based on the actual outdoor measurements during the analysis period. Details on the measurement methods and gathered parameters can be found in chapter 4. The following outdoor measurements were integrated into the customized epw file: Outdoor Temperature (°C), Outdoor Humidity (Relative Humidity (%)), Precipitation (mm/hr), Diffuse Radiation (W/m²), Direct Radiation (W/m²), Global Radiation (W/m²), Wind Speed (m/s), and Wind Direction (degrees (°) from North). The customized epw file was generated using Python. A visual analysis of the environmental conditions applied in this simulation is shown in Figure 8.4.

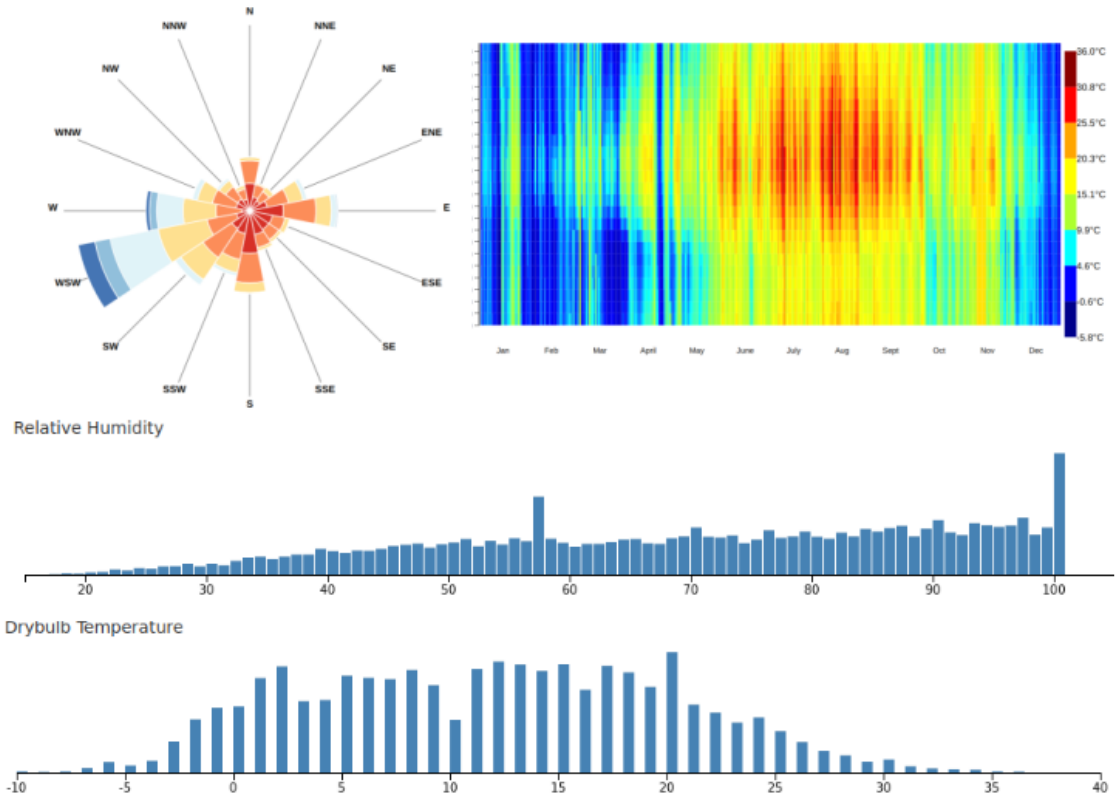


Figure 8.4: Visualization of the epw file using epwvis (https://mdahlhausen.github.io/epwvis/). Top Left: Wind speed and direction frequency Top Right: Temperature distribution by hour over a year. Mid: Histogram of relative humidity. Bottom: Histogram of drybulb temperature.

8.4 Model Validation

This section describes the validation of the simulation model and presents the corresponding validation results.

8.4.1 Procedure

The thermal simulation model will be calibrated based on indoor air temperature and CO_2 concentration measurements, and the simulation model outputs will be compared to the actual measured values. The Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) (Equations 8.1 and 8.2) will be used as validation metrics, calculated separately for CO_2 and indoor air temperature.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8.1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8.2}$$

The initial 1000-hour period of the simulation will be excluded from the calibration process due to transient responses of the simulation model resulting from the preheating processes of the thermal mass. These responses caused significant deviations between the model and the observed values that could not be explained by the model parameters. Additionally, the month of August was removed from the validation data since HVAC and occupancy levels were reduced due to holidays, resulting in deviations that could not be accurately modeled.

8.4.2 Validation results

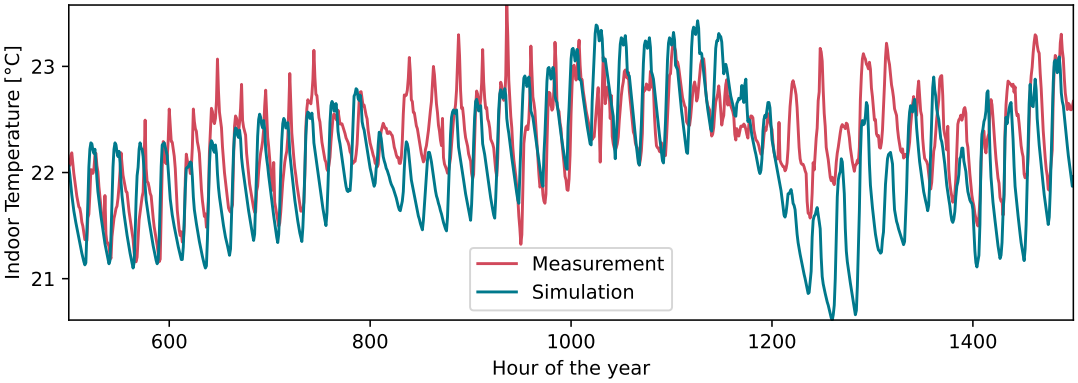


Figure 8.5: Exemplary selection from the validation data for indoor air temperature

Indoor Air temperature The VDI 6020 "Anforderungen an thermisch-energetische Rechenverfahren zur Gebäude- und Anlagensimulation" (Requirements for thermal energy calculation methods for building and system simulation) defines a maximum MAE of 1.0 K for calibrated simulation models. The achieved MAE of 0.49 K is significantly below this threshold. Furthermore, the RMSE (0.624 K) is only slightly higher than the MAE, indicating a low level of deviation between the model and the measured values.

CO_2 concentration There are no established guidelines for validating CO_2 indoor air concentration metrics. However, considering the MAE (15.5 ppm) and RMSE (21.8 ppm) in relation to the total variance of indoor CO_2 concentration (225 ppm), the MAE represents 6.7% of the variance and the RMSE represents 9.6% of the variance.

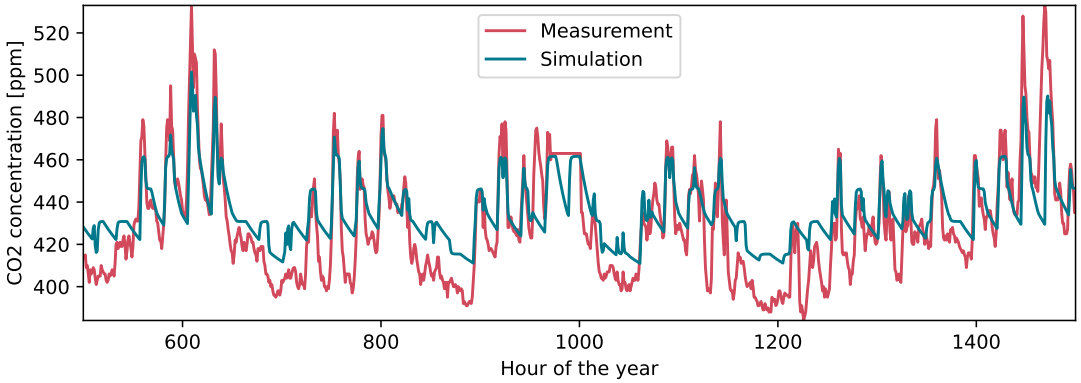


Figure 8.6: Exemplary selection from the validation data for CO_2

Table 8.1: Validation results of the thermal simulation model for indoor air temperature and CO_2 concentration

Metric	MAE	RMSE
Indoor Temperature	0.499	0.624
CO_2 Concentration	15.5	21.8

In summary, a calibrated thermal simulation model was generated by modeling the building parts and geometry according to the building plans. The building operation and occupancy patterns were calibrated using real measurements and a customized weather file. The validation results for indoor air temperature and CO_2 concentration showed that the simulation model accurately models indoor conditions. Therefore, the model is suitable for studying the integration of virtual indoor air pollutant sensors in demand-controlled ventilation.

8.5 Control Strategies

To evaluate the performance of demand-controlled ventilation using virtual indoor air pollutant sensors, a comparison of various control strategies will be made, evaluating their air treatment and transportation energy consumption. For this, the six major operation modes defined in DIN EN 16798-3 will be used as a basis for the control strategies. The control strategies are detailed in the following section, and a summary of the control strategies is given in Table 8.2.

Table 8.2: Description of the IDA Control Categories as defined in DIN EN 16798-3

Category	Description
IDA – C 1	The system is operated constantly.
IDA – C 2	The system is manually controlled **(not considered)**.
IDA – C 3	The system is operated according to a predetermined schedule.
IDA – C 4	The system is operated based on the presence of people (light switches, infrared sensors, etc.).
IDA – C 5	The system is operated in steps depending on the number of people in the room.
IDA – C 6	The system is controlled by sensors that measure indoor air quality parameters or adjusted criteria (e.g., CO_2, mixed gas, humidity, or VOC sensors).

8.5.1 IDA1: Continuous control

In the IDA1 control strategy, the ventilation system is operated constantly. The supply air volume is set to a constant air change rate of 1.6 air changes per hour (as currently operated).

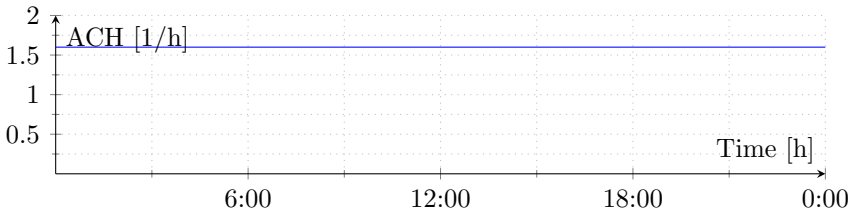


Figure 8.7: Exemplary volume flow for IDA1 control strategy

8.5.2 IDA3: Scheduled control

The IDA3 control strategy operates the ventilation system based on a predetermined schedule. IDA3 represents the current operating mode in the open office. In this specific case, the ventilation system is activated between 5:15 am and 8 pm on workdays, providing a constant supply air volume of 1.6 air changes per hour.

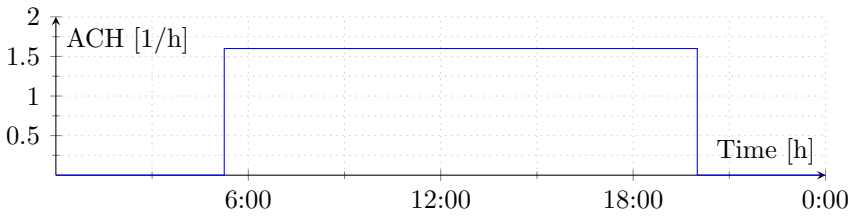


Figure 8.8: Exemplary volume flow for IDA3 control strategy

8.5.3 IDA4: Occupancy-based control

Control strategy IDA4 is based on occupancy signals. When one or more occupants are present, the ventilation system is activated and provides a constant supply of 1.6 air changes per hour.

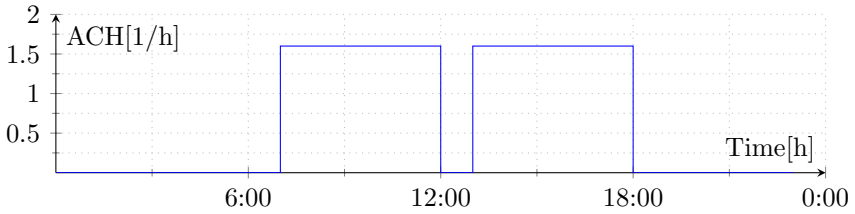


Figure 8.9: Exemplary volume flow for IDA4 control strategy

8.5.4 IDA5: Adaptive Occupancy-based control

The ventilation system in control strategy IDA5 operates based on occupancy and adjusts the air supply volume according to the number of occupants. The maximum air supply of 1.6 air changes per hour is reduced proportionally to the maximum capacity of 35 occupants.

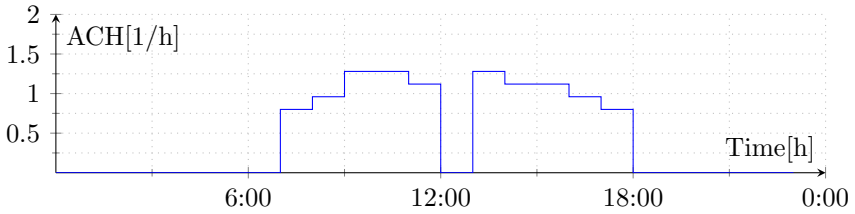


Figure 8.10: Exemplary volume flow for IDA5 control strategy

8.5.5 IDA6: Pollutant-based control

The IDA6 control strategy is based on the evaluation of indoor air pollution in the room, specifically CO_2, VOC, and $PM_{2.5}$ concentrations. Control signals are calculated for each

pollutant every minute, comparing them to health-relevant thresholds of 1000 ppm CO_2, VOC index of 250, and 10 $\mu g/m^3$ $PM_{2.5}$, which are derived from EU regulations and WHO guidelines (see chapter Theory). If any of the control signals exceed the threshold, the air supply volume is set to 1.6 air changes per hour. In this case, a simple on/off hysteresis controller with a deadband of 200 ppm for CO_2, 50 IAQI for VOC, and 2 $\mu g/m^3$ for $PM_{2.5}$ is implemented. While an implementation using PI or PID controllers is also possible, the simple on/off controller is sufficient for the purposes of this thesis. If none of the control signals exceed the threshold, the air supply volume is set to 0. If one or more control signals exceed the threshold, the supply air volume is set to 1.6 air changes per hour.

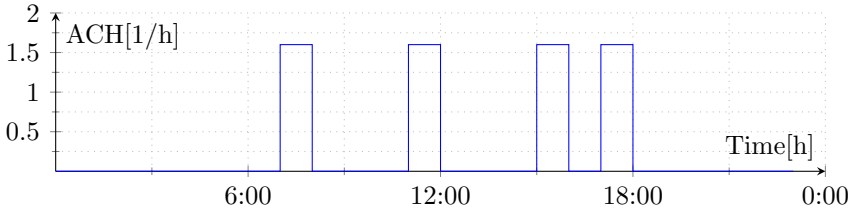


Figure 8.11: Exemplary volume flow for IDA6 control strategy

8.5.6 ID6: Pollutant-based control scenarios

In addition to the selected control strategies, three simulation scenarios have been defined for IDA6 to explore the effects of varying occupancy and pollution patterns on the air treatment and transportation energy consumption of the control strategy implementing virtual indoor air pollutant sensors. The scenarios are defined as follows:
S0: Base case. Scenario S0 represents the base case as presented above. The boundary conditions are the same as those explored for the other ventilation control strategies IDA1 - IDA5.
S1: Maximum occupancy. During the observation period, the room occupancy was far below the maximum at all times. Scenario S1 explores the effects of maximum occupancy in the room on the air treatment and transportation energy consumption of control strategy IDA6. Maximum occupancy is applied when people are present.
S2: Increased pollution. Scenario S2 explores the effect of increased non-anthropogenic (VOC and $PM_{2.5}$) pollution due to internal or external sources. In this scenario, the originally measured values are increased by a factor, leading to more frequent exceedances of pollutant thresholds and therefore an increase in ventilation time.
S3: Maximum occupancy and increased pollution. Scenario S3 combines scenarios S1 and S2, increasing both the occupancy and indoor air pollution.

8.6 Results

This section presents the results of case study 1, which tests demand controlled ventilation strategies in a validated simulation model that includes the integration of virtual indoor air pollutant sensors. The focus of the evaluation is on air treatment and transportation energy consumption. The methods used to develop the models are described in the previous section. .

8.6.1 IDA 1

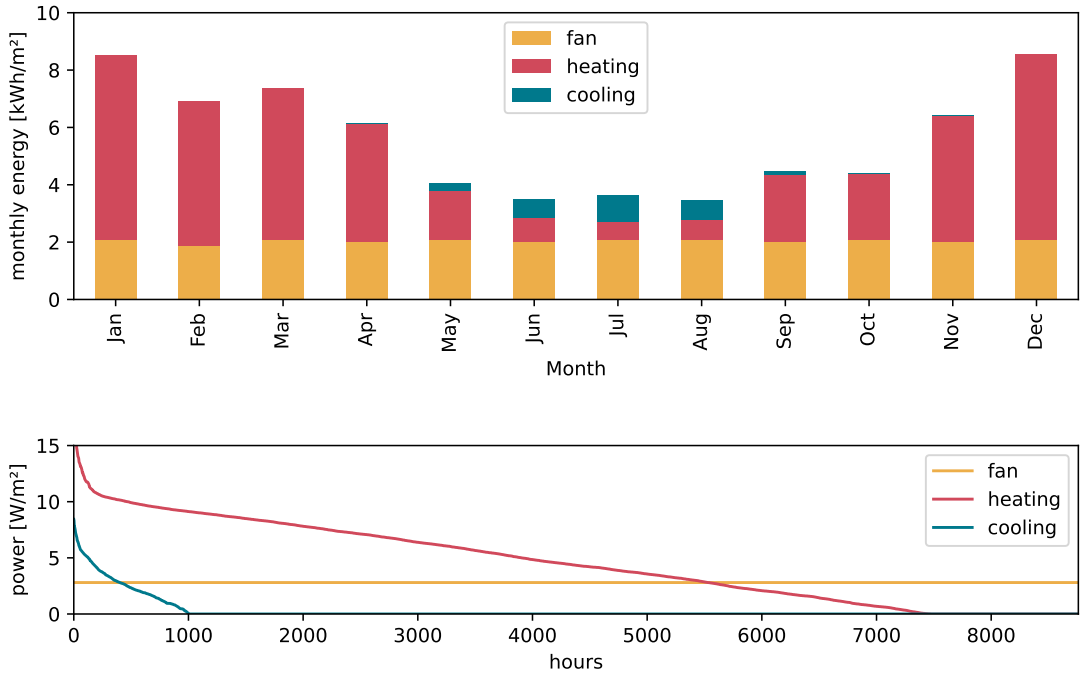


Figure 8.12: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA1

The energy consumption of control strategy IDA1 in the office room is examined in this section. The Figure displays the monthly air treatment and transportation energy consumption, distinguishing between ventilation, heating, and cooling. Ventilation includes the electricity demand for fan operation, heating refers to the energy required by heating coils, and cooling represents the energy demand for cooling coils. The figure demonstrates that the greatest overall energy consumption is attributed to preheating the air to the required supply air temperature, followed by fan operation. Cooling energy is only required during the months of May to August. Heating demand is highest in the winter months, while ventilation energy demand remains constant throughout the

observational period. The annual load duration curve in the Figure provides a visual representation of the continuous operation of the ventilation system throughout the observational period. Heating is active for over 7000 hours, while cooling is active for 1000 hours. The peak power demand for heating is 15 W/m², while cooling requires 7.5 W/m². The power demand for ventilation remains constant at 4 W/m².

8.6.2 IDA 3

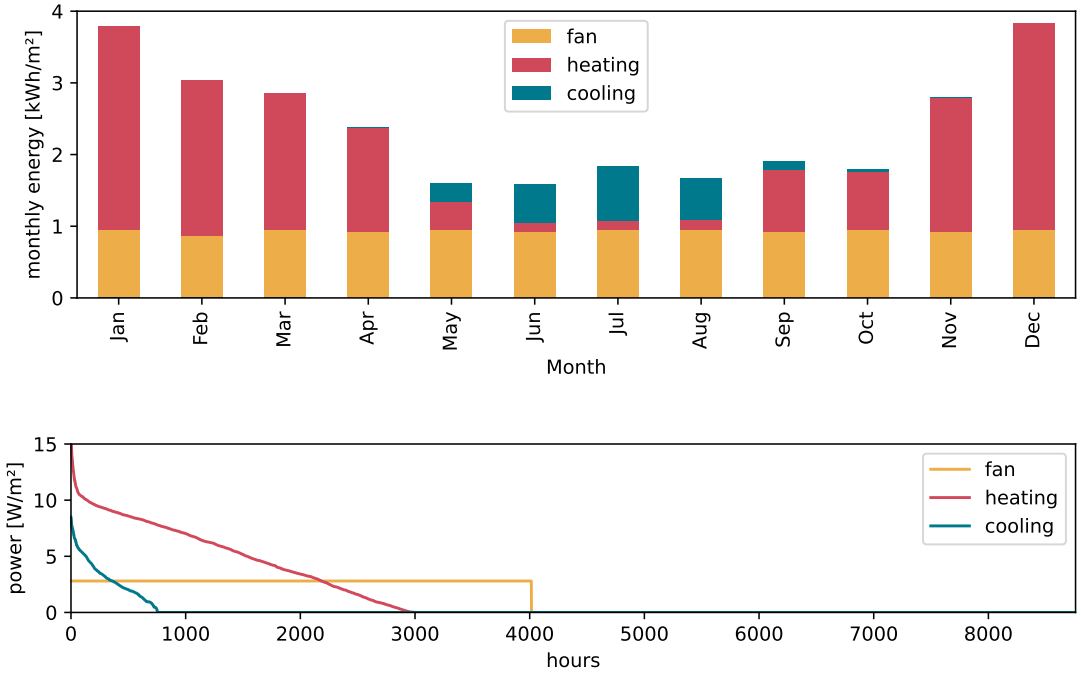


Figure 8.13: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA3

The Figure presents the monthly energy consumption of control strategy IDA3 (on-off schedule), which demonstrates a reduction in energy consumption compared to IDA1 for each month. Similar to IDA1, heating energy demand accounts for the majority of the overall energy consumption, followed by fan operation. However, IDA3 requires a higher proportion of cooling energy during the months of May to September. This is due to the absence of a nighttime cooling effect from the building's thermal mass, resulting in increased daytime cooling demand. Implementing an additional nighttime cooling schedule during the summer months could effectively reduce cooling power during these periods. The annual load duration curve for IDA3 reflects a reduced operation time of approximately 4000 hours compared to IDA1. The constant volume flow of the ventilation system is evident from the curve, as the power consumption remains constant.

The peak power demand for heating and cooling in IDA3 is the same as in IDA1, at 15 W/m² and 7.5 W/m², respectively.

8.6.3 IDA 4

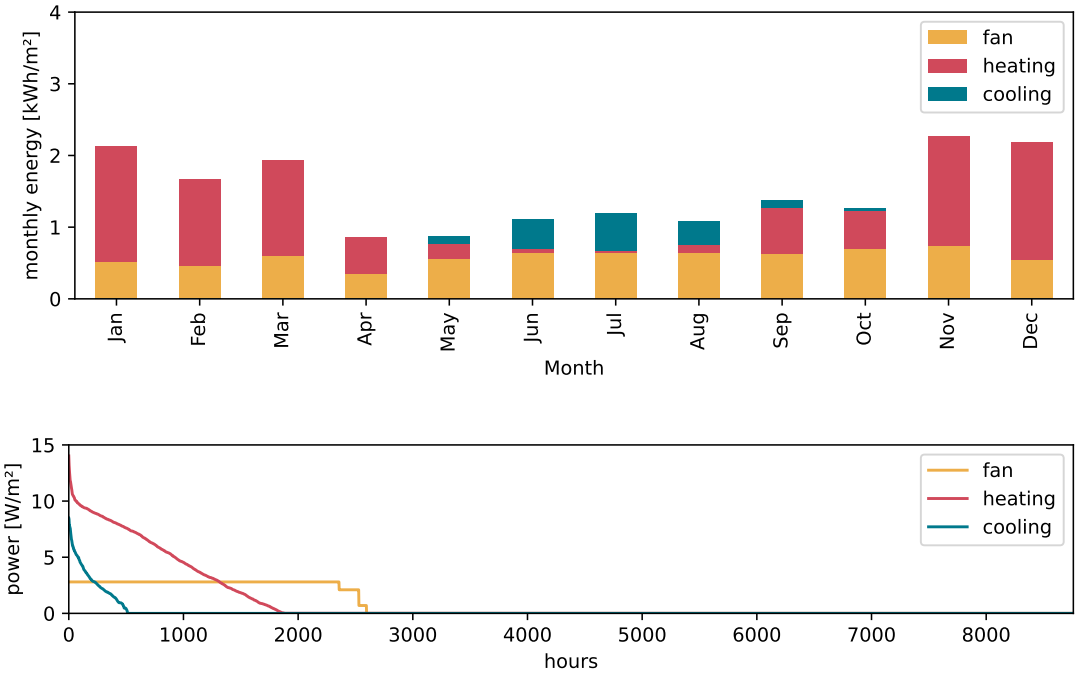


Figure 8.14: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA4

The monthly energy consumption of control strategy IDA4, which employs on-off presence control, demonstrates a reduced energy consumption compared to the previous strategies IDA1 and IDA3. The dominant factors contributing to energy consumption are heating and ventilation. However, in the summer months, cooling can account for up to half of the energy consumption. Since presence varies across the months, the demand for ventilation energy is also variable, unlike in the IDA1 and IDA3 control strategies. The annual load duration curve reveals that the ventilation system is active for approximately 2500 hours. The peak power consumption for heating is lower than in the ventilation control strategies IDA1 and IDA3, at 13 W/m². The peak cooling power is the same as in the IDA1 and IDA3 control strategies. Although air supply is a binary (on/off) process, a graduated specific power pattern can still be observed in the annual load duration curve. This can be attributed to the hourly mean value, where the ventilation system may cycle on and off multiple times.

8.6.4 IDA 5

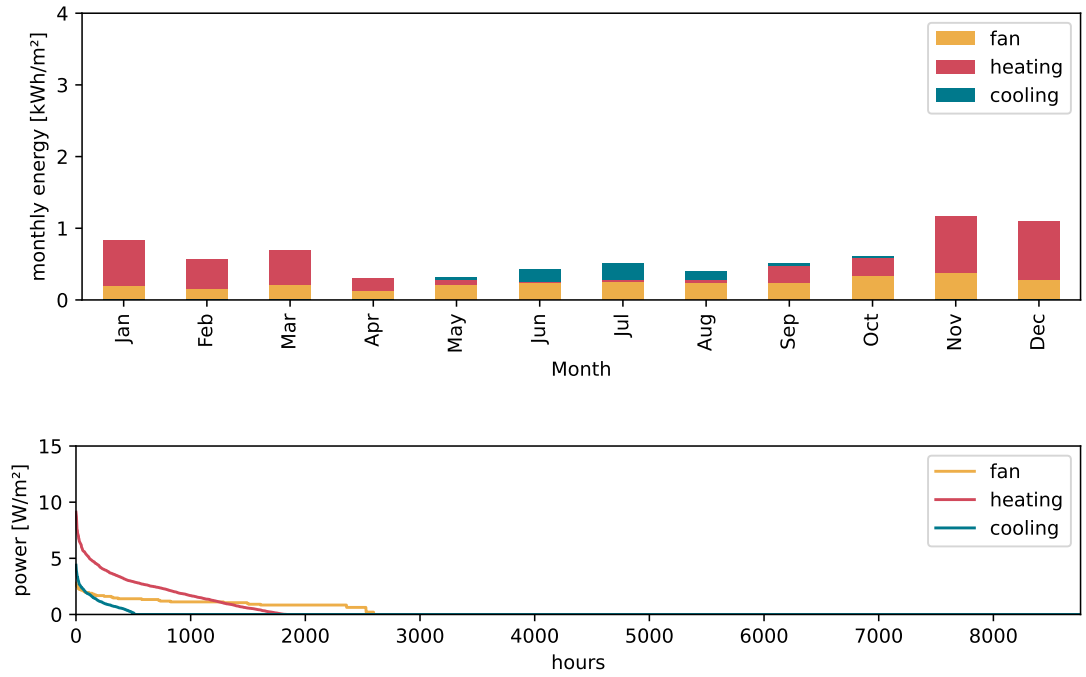


Figure 8.15: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA5

The monthly energy consumption of control strategy IDA5, which employs occupancy-based control, reduces the energy demand of each month by more than half compared to the IDA4 control strategy. This reduction is achieved by controlling the supply air volume based on the number of people present in the room. Heating accounts for about two-thirds of the total energy demand, while ventilation accounts for one-third. Cooling energy demand is negligible in June to August. The annual load duration curve shows a continuous ventilation curve since the supply air volume is variably controlled based on the number of people present. The ventilation system is active for approximately 2500 hours throughout the year. The peak energy demand for heating and cooling is reduced compared to control strategies IDA1 - IDA4, as a result of lower supply air volumes during occupancy, as the room was never fully occupied. The peak heating power is 7.5 W/m² and the peak cooling power is 4 W/m².

8.6.5 IDA 6

The figures depict the monthly energy consumption of control strategy IDA6, which incorporates virtual indoor air pollutant sensors to activate ventilation based on health-relevant pollutant thresholds. During January, February, and May through October,

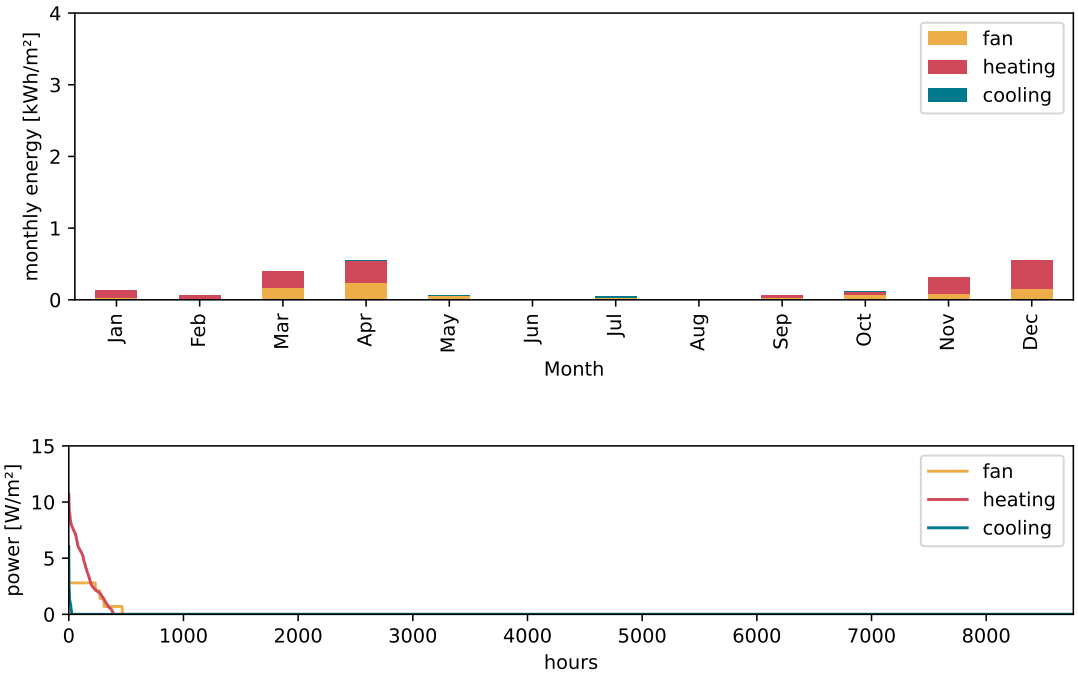


Figure 8.16: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA6

no energy is consumed. Energy consumption in the remaining months is attributed to heating and ventilation, with cooling energy demand being negligible in control strategy IDA6. The annual load duration curve illustrates that the ventilation system is active for only 500 hours per year. The peak heating energy demand slightly exceeds that of control strategy IDA5 at 9 W/m². Cooling can be disregarded.

8.6.6 IDA 6 scenario analysis

In order to assess the energy consumption of the ventilation control strategy IDA 6, which incorporates the indoor air pollutant virtual sensing model, three scenarios were investigated alongside the base case. These scenarios are as follows: S0, which represents the base case; S1, which assumes maximum occupancy; S2, which involves an increase in pollution; and S3, which combines maximum occupancy and increased pollution.
The results of the scenario analysis are presented in the Figure. The bar plots illustrate the total annual energy consumption, with a breakdown of heating, cooling, and fan energy demand. The base scenario S0 exhibits the lowest overall power consumption. When maximum occupancy is assumed (S1), the total power consumption nearly doubles due to increased ventilation requirements and longer runtimes for discharging anthropogenic air pollutants. However, it should be noted that the results of scenario S1 are still

significantly lower than those of the ventilation strategy IDA5. In scenario S2, where non-anthropogenic air pollutants are increased, the annual energy consumption of the ventilation system increases more than fivefold. This is because pollutant thresholds are more frequently exceeded, leading to longer ventilation runtimes. Scenario S3 shows only a minor increase in total energy consumption compared to scenario S2, suggesting that increased occupant levels have a limited impact. However, it is noted that the simulation does not consider the removal of non-anthropogenic pollutants through increased air supply. Therefore, scenarios S2 and S3 may overestimate the energy consumption as the pollutant discharge of the ventilation system is not accounted for.

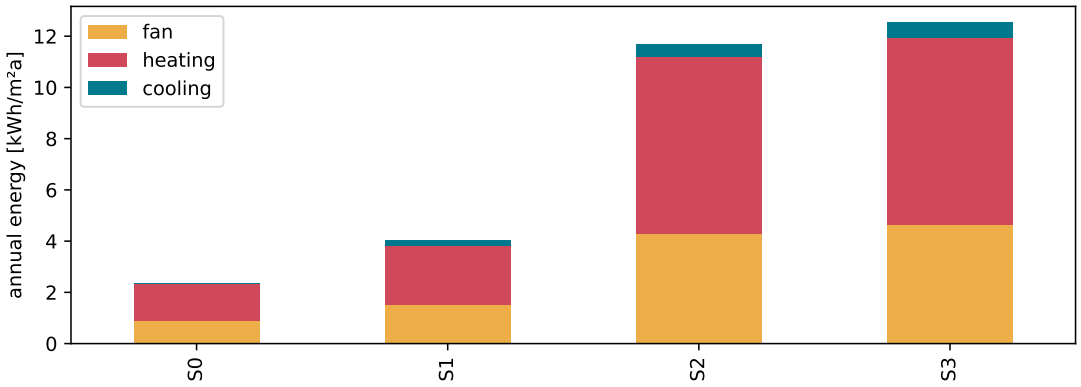


Figure 8.17: Annual air treatment and transportation energy consumption in control strategies IDA6 in scenarios S0-S3

8.6.7 Comparison

The following section compares the annual energy consumption among all ventilation control strategies. The Figure displays the annual energy consumption of the ventilation system, with a breakdown of heating, cooling, and fan energy. The total energy consumption decreases as more complex ventilation control strategies are employed, allowing for demand-driven ventilation control rather than a constant supply. The greatest reduction in energy consumption is achieved by implementing an indoor air pollutant-based control strategy using the virtual sensing model (IDA6). IDA6 consumes only 3.2% of the air treatment and transportation energy required by IDA1 and is able to reduce the energy consumption of the usually in demand controlled ventilation employed occupancy-controlled ventilation strategy IDA5 by 72%. Additionally, the operating hours of IDA6 are significantly reduced, resulting in lower maintenance costs. The simulation also indicates that cooling is not necessary in the optimized ventilation strategy, further decreasing overall costs.

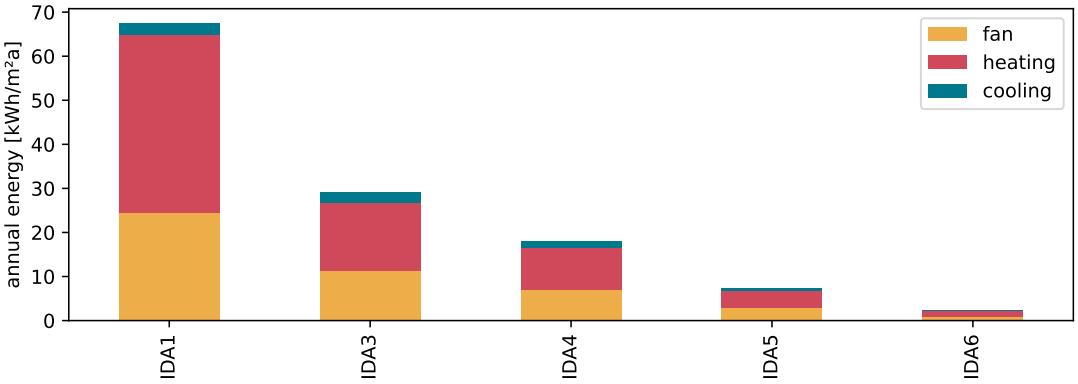


Figure 8.18: Annual air treatment and transportation energy consumption in control strategies IDA1-IDA6

9 Case Study 2 - Classroom

In case study 2 a decentralized ventilation unit in the Classroom is monitored and controlled, allowing for on-site testing of various control strategies and monitoring their impact on indoor air pollutants and air treatment and transportation energy consumption. To expand the testing to all control strategies previously described in case study 1, a co-simulation is employed that was validated on the monitoring data, allowing to estimate the air treatment and transportation energy consumption of the ventilation unit under all operation modes. The results of the case study are presented in the following chapter.

9.1 Test Space

The second case study takes place in the Classroom that was previously used to evaluate the transferability of the virtual sensing model. The configuration of the ventilation unit, as well as the measurements conducted for evaluation, are described below. A detailed description of the building, classroom, and measurements can be found in the previous chapters of this dissertation.

9.1.1 Decentralized Ventilation Unit

The decentralized ventilation unit installed in the Classroom is used as a testing environment for demand-controlled ventilation strategies. The specific unit used is the Trox SCHOOLAIR-V-HV-EH model, in dual configuration. To achieve the required volume flow rates, two identical units were installed. The units are controlled in parallel using the master-slave principle.
The specifications of the dual unit Trox SCHOOLAIR-V-HV-EH are presented in the table.
For control, the decentralized ventilation unit utilizes integrated sensor technology. The

Table 9.1: Specifications of the dual unit Trox SCHOOLAIR-V-HV-EH

Function	**Unit Component**	**Specification**
Heat exchange	Rotary heat exchanger	75% efficiency
Heating	Electric heating coil	1500 W heating power
Filtration	Air filter filter class	EPM1 65%, exhaust air ISO Coarse 50%
Air flow control	Maximum volume flow	400 m^3/h

unit incorporates the sensor systems as shown in table 9.2.

The ventilation unit is installed in default with a standard control configuration specified

Table 9.2: Trox SCHOOLAIR-V-HV-EH integrated sensing

Sensor System	Function	Unit
CO_2	Measures CO_2 concentration in the exhaust air for demand controlled ventilation	ppm
Supply air temp.	Controls the heat exchanger and the heating coil	°C
Outdoor air temp.	Controls the heat exchanger and the heating coil	°C

by the manufacturer. The control is optimized for classrooms and includes a predefined schedule that activates the ventilation unit between 8:00 a.m. and 6:00 p.m. The supply air volume is controlled based on CO_2 levels. Additionally, the heating coil is activated when the supply air temperature falls below 20 °C. Users also have the option to open windows for natural ventilation. To evaluate the performance of different control strategies, the default control configuration of the ventilation unit was replaced by custom control algorithms. This was achieved by utilizing the read/write interface of the unit, which provided access to all control variables.

9.1.2 Measurements

In addition to the measurements from the ventilation unit, a monitoring system was installed to record indoor comfort, air quality, use of the classroom, outdoor climate, and power consumption of the ventilation unit. The ventilation unit was integrated into the monitoring system via available interfaces to gather operation data and control its operation. This interface enables the readout of current operating parameters and room climate measured variables, as well as the control of individual components of the unit. The figure shows the schematic measurement setup in the Classroom.

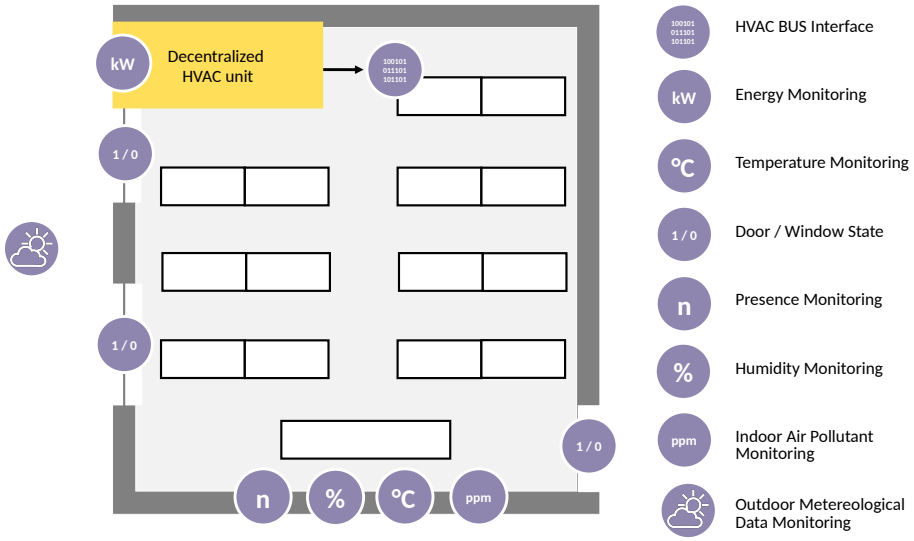


Figure 9.1: Measurement setup

9.2 Testing framework

The testing of ventilation control strategies is performed using a hybrid approach that combines control and monitoring of the ventilation unit with a validated co-simulation method. The testing framework is illustrated in Figure 9.2. Real device measurements were performed for control strategies IDA 3 and IDA 4, while the co-simulation method was applied for IDA 1, 5, and 6.

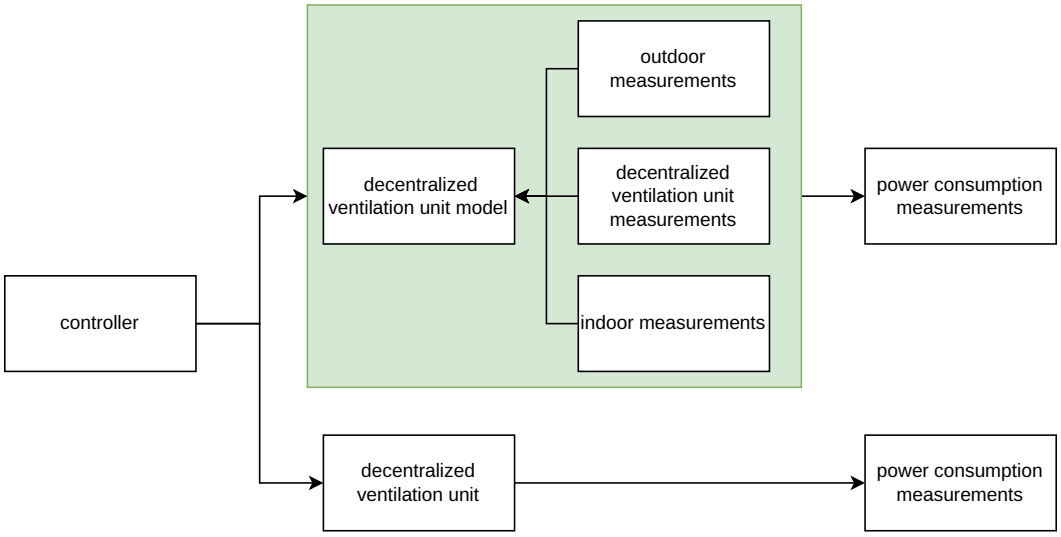


Figure 9.2: Testing framework

9.3 Co-Simulation Setup

A Python-based co-simulation is developed to physically model the decentralized ventilation unit. The focus is on the main components of the unit, including heat recovery, heating coils, and fans. The co-simulation is calibrated using operation data from the ventilation unit and validated against measurements from the monitoring system. Figure 9.3 provides a graphical representation of the co-simulation. The following sections

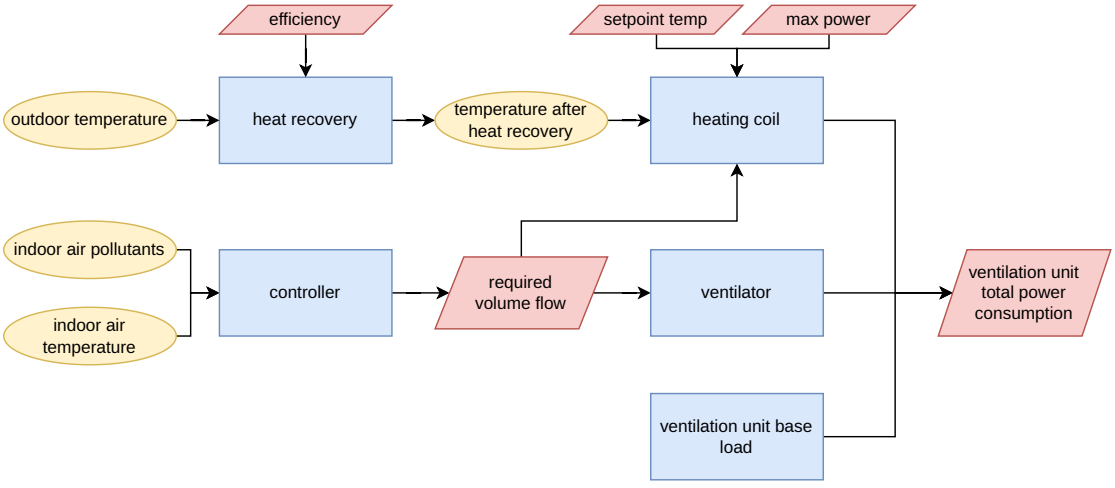


Figure 9.3: Co-simulation method

present the equations that define the main components of the co-simulation.

Heat Recovery

$$t_{wrg} = (t_{in} - t_{out}) \cdot n_{wrg} + t_{out} \tag{9.1}$$

This equation calculates the resulting temperature (t_{wrg}) after heat recovery. Here, t_{in} represents the initial temperature, t_{out} represents the temperature after heat recovery, and n_{wrg} represents the heat recovery efficiency.

Heating Coil

$$P = \rho \cdot x \cdot c_p \cdot \Delta T \tag{9.2}$$

This equation calculates the power requirements (P) to heat air with a volume flow rate (x) from an inlet temperature (t_{in}) to an outlet temperature (t_{out}) using the specific heat capacity (c_p). The resulting power requirements are measured in Watts. Here, ρ represents the density of air in kg/m^3 at standard temperature and pressure, ΔT represents the temperature difference in degrees Celsius (°C), and t_{in} and t_{out} represent the temperature of air at the inlet and outlet, respectively.

Fan

$$P = q \cdot p_s \tag{9.3}$$

This equation calculates the power consumption (P) of a fan. The resulting power consumption is measured in Watts. Here, q represents the volume flow rate of air in cubic meters per second (m^3/s), and p_s represents the specific power consumption in cubic meters per Watt per second (m^3/Ws). The value of p_s is determined to be 0.00139 m^3/Ws based on previous measurements.

Total Power Consumption

$$P_{tot} = P_{vent} + P_{heat} + P_{standby} \tag{9.4}$$

This equation calculates the total power consumption (P_{tot}) of the HVAC system. Here, P_{vent} represents the power consumption of the fan in Watts, P_{heat} represents the power requirements to heat air in Watts, and $P_{standby}$ represents the standby power consumption in Watts. The value of $P_{standby}$ is determined to be 25 W based on previous measurements.

9.4 Co-Simulation Validation

The model is validated against measurement and operation data at known operating states to ensure validity. Validation is performed across a wide range of operating conditions. Evaluation metrics, namely Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), are calculated. The validation test is conducted for both the heating and cooling period. Figure 9.4 and 9.5 present a 24-hour excerpt of the heating and cooling period, comparing the co-simulation and monitoring results. During the heating period, the MAE metric for the power consumption of the device for the entire testing period is **14.1** W, with a corresponding RMSE of **66.2** W. Figure 9.4 depicts the comparison between

the model and measurements during the heating period. The figure demonstrates that the model accurately represents the power consumption during standby. However, during device activity, the measurements exhibit a more transient response with an initial power peak that is not captured in the model. Nevertheless, the model and measurements align closely after this peak. The cause of this peak power consumption is unclear. Nonetheless, measurements during the cooling period indicate that it is related to the heating register, implying a transient response during the preheating process. However, due to the short-term nature of this deviation, it is not considered in the model. Figure

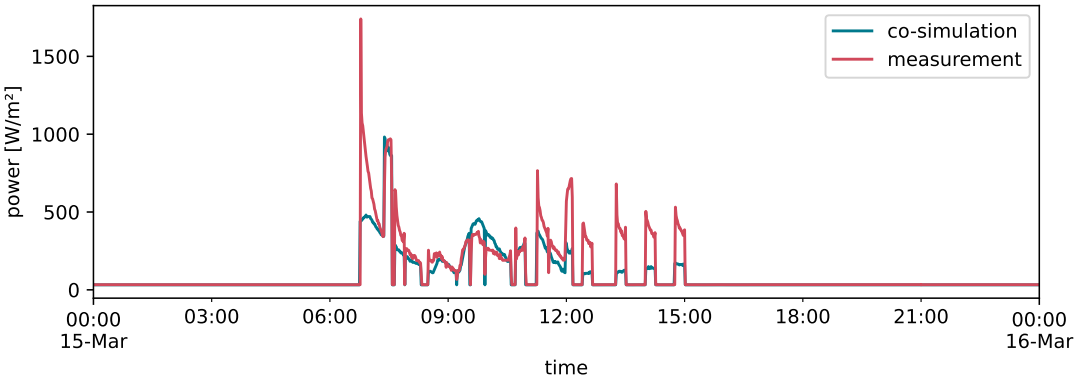


Figure 9.4: Comparison of Model and Measurements during the heating period

9.5 illustrates the co-simulation and measured values for an exemplary 24-hour timeframe during the cooling period. The model and measurements exhibit a very good fit with minimal differences. Thus, it can be concluded that the model accurately represents fan power consumption across different ventilation stages.

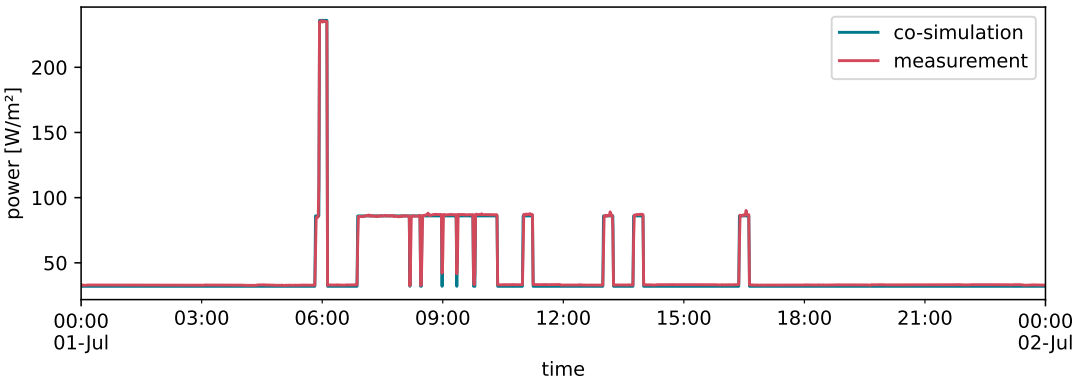


Figure 9.5: Comparison of Model and Measurements during the Summer

Table 9.3 provides monthly performance metrics for model predictions during the heating and summer period. Higher errors are observed during the winter months due to

Table 9.3: Monthly performance metrics for model predictions

Month	**MAE (W)**	**RMSE (W)**
January	23.0	80.6
February	54.9	142.4
March	28.4	90.6
April	10.2	47.1
May	1.5	4.7
June	1.1	2.6
July	1.4	4.2
August	0.9	1.5
September	1.1	4.0
October	1.2	2.7
November	22.3	87.6
December	30.8	101.3

transient responses of the heating register after activation, resulting in increased MAE and RMSE values. The peak MAE of **54.9** W in February, although low compared to the variability in power consumption of 2000 W, can still be attributed to the transient response. In contrast, the very low MAE and RMSE values during the summer months indicate accurate prediction of the power consumption of the ventilation unit with minimal deviations. These deviations during the winter months are primarily caused by the transient response of the heating register during preheating. In summary, the model accurately predicts the power consumption of the ventilation unit, as demonstrated by the testing of various control strategies. This accuracy is achieved by physically modeling the processes in the ventilation unit and calibrating them with factors based on the device's measurements.

9.5 Control strategies

This section presents the control strategies explored in this case study. Analog to case study 1, the DIN EN 16798-3 control strategies will be used as a baseline for comparison, with IDA 6 (pollutant-based control) incorporating the virtual indoor air pollutant sensor.

9.5.1 IDA1: Continuous control

In the IDA1 control strategy, the ventilation system operates continuously. The supply air volume is set to 30 m^3 per hour per person, assuming an occupation of 25 pupils and one teacher per classroom. Thus, a total of 800 m^3/h of supply air will be delivered.

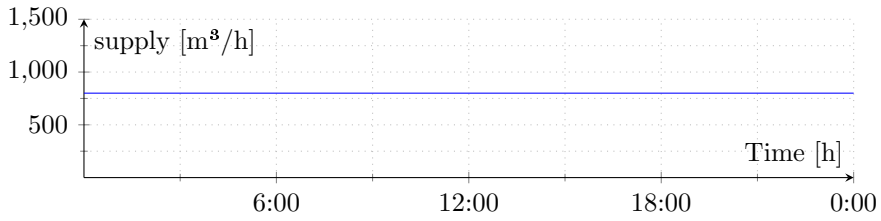


Figure 9.6: Exemplary volume flow for IDA1 control strategy

9.5.2 IDA3: Scheduled control

In the IDA3 control strategy, the ventilation system operates between 8:00 a.m. and 6:00 p.m. The supply air volume is set to 30 m^3 per hour per person, assuming an occupation of 25 pupils and one teacher per classroom, resulting in a total of 800 m^3/h of supply air delivered per hour.

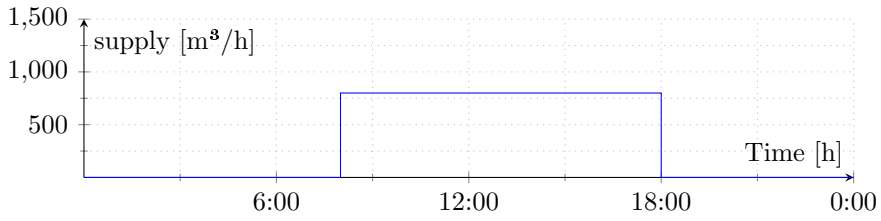


Figure 9.7: Exemplary volume flow for IDA3 control strategy

9.5.3 IDA4: Occupancy-based control

In the IDA4 strategy, the ventilation system is only operated during occupancy. To ensure the removal of remaining pollutants after occupants leave the room, an activity decay time of 15 minutes is implemented. The supply air volume is set to 30 m^3 per hour and person, assuming an occupancy of 25 pupils and one teacher per classroom. Therefore, a total of 800 m^3/h of supply air will be delivered during occupancy.

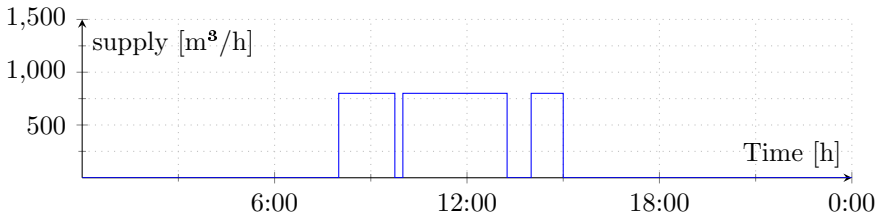


Figure 9.8: Exemplary volume flow for IDA4 control strategy

9.5.4 IDA5: Adaptive Occupancy-based control

In the control strategy IDA5, the ventilation system is only operated during occupancy with a variable supply air volume based on the number of occupants in the room. The supply air volume ranges from 400 m³ per hour to 1200 m³ per hour.

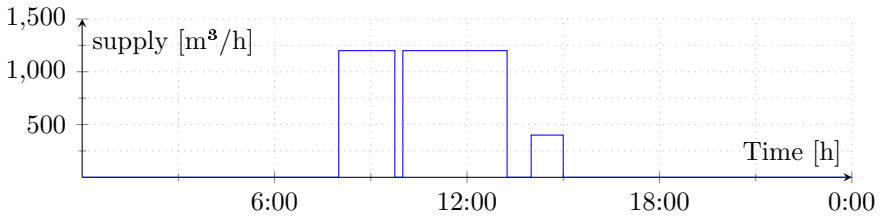


Figure 9.9: Exemplary volume flow for IDA5 control strategy

9.5.5 IDA6: Pollutant-based control

In the control strategy IDA6, the ventilation system is operated based on the indoor air pollution in the room, evaluating the concentration of CO_2, VOC, and $PM_{2.5}$ in the indoor air as measured by the virtual indoor air pollutant sensor. A control signal is calculated for each pollutant every minute to determine if health-relevant thresholds of 1000 ppm CO_2, a VOC index of 250, and 10 $\mu g/m^3$ of fine $PM_{2.5}$ are exceeded. The supply air volume is adjusted according to the number of active control signals and their respective exceedances. The supply air volume ranges from 400 m³ per hour to 1200 m³ per hour.

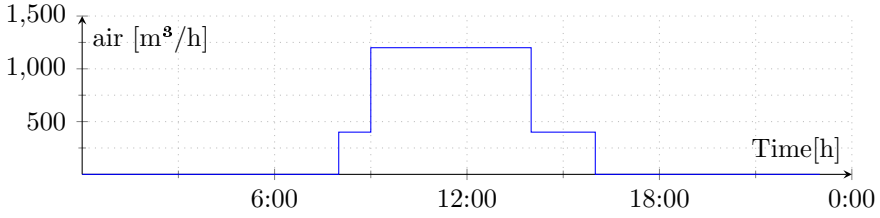


Figure 9.10: Exemplary volume flow for IDA6 control strategy

9.6 Results

In this section, the air treatment and transportation energy demand will be compared for all ventilation control strategies. The energy consumption will be differentiated into heating energy and fan energy. Cooling energy will not be considered since the decentralized ventilation unit does not have a cooling register installed.

9.6.1 IDA 1

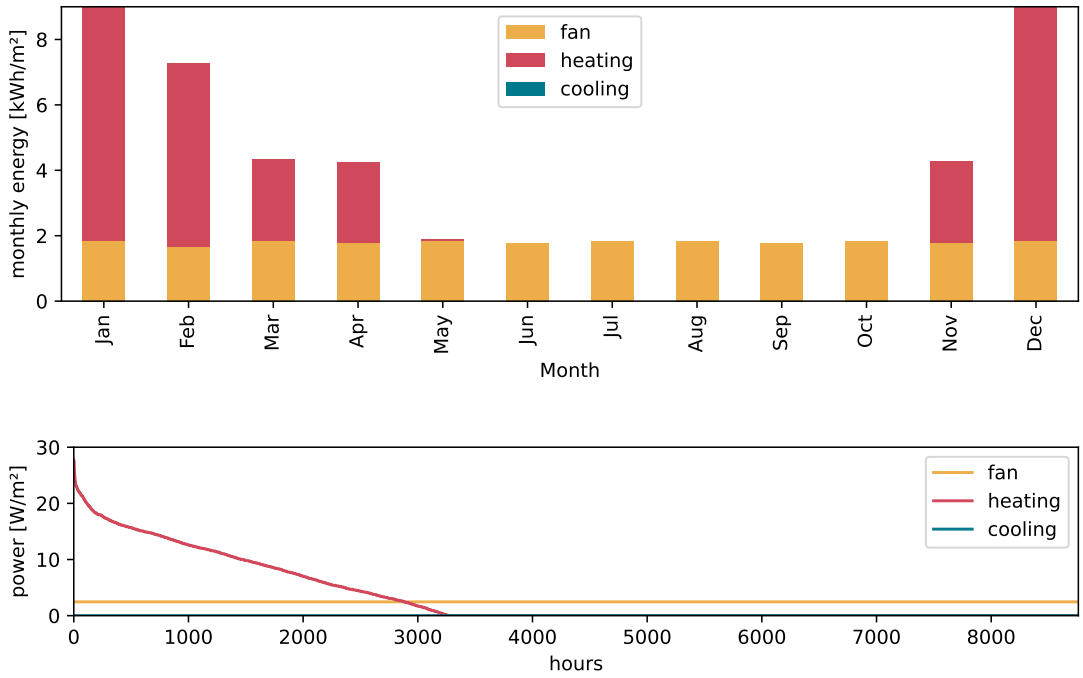


Figure 9.11: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA1

The figure illustrates the monthly energy consumption of the ventilation control strategy IDA1 (always on). 40% of the annual energy demand is attributed to the fans, while 60% is used for heating the supply air. The monthly energy demand is highest in December and January due to low outside air temperatures, resulting in high preheating requirements. From May to October, no supplemental heating is necessary, and the power consumption of the fans becomes the determining factor. The annual load duration curve demonstrates that the fans operate throughout the entire year, while heating is only required for approximately 3000 hours with a peak heating power of 24 W/m².

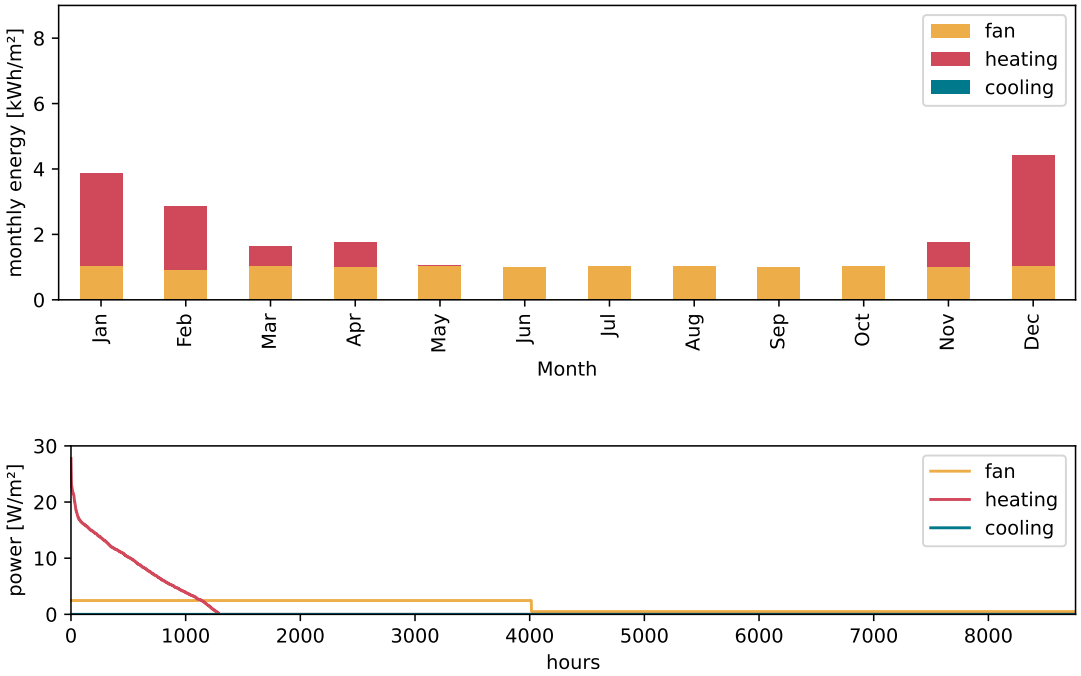


Figure 9.12: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA3

9.6.2 IDA 3

The figure illustrates the monthly energy consumption associated with the ventilation control strategy IDA3 (scheduled on-off). The total energy demand in IDA3 is less than half of that in control strategy IDA1, primarily due to reduced runtimes. Annually, IDA3 results in a 57% reduction in energy consumption compared to control strategy IDA1. The annual energy demand is divided almost equally between heating and fan power consumption.
The annual load duration curve reveals that the ventilation unit operates for 4000 hours per year, with heating active during 1200 hours. The peak heating power demand is 21 W/m².

9.6.3 IDA 4

The figure illustrates the monthly energy consumption associated with the ventilation control strategy IDA4 (presence based on-off). Control strategy IDA4 reduces the annual energy consumption of IDA3 by 56% and that of IDA1 by 81%. Compared to the ventilation control strategies IDA1 and IDA3, fan energy consumption dominates the total energy consumption annually. The annual load duration curve reveals that the

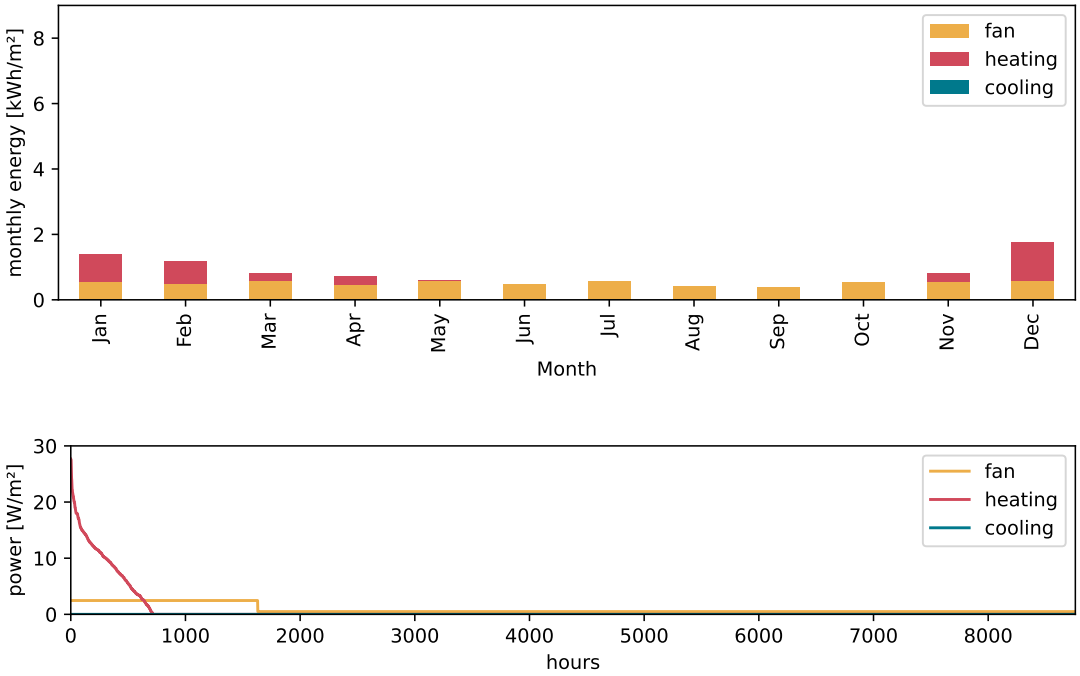


Figure 9.13: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA4

ventilation unit operates for 1600 hours per year, with heating active during approximately 750 hours. The peak heating power demand is 21 W/m².

9.6.4 IDA 5

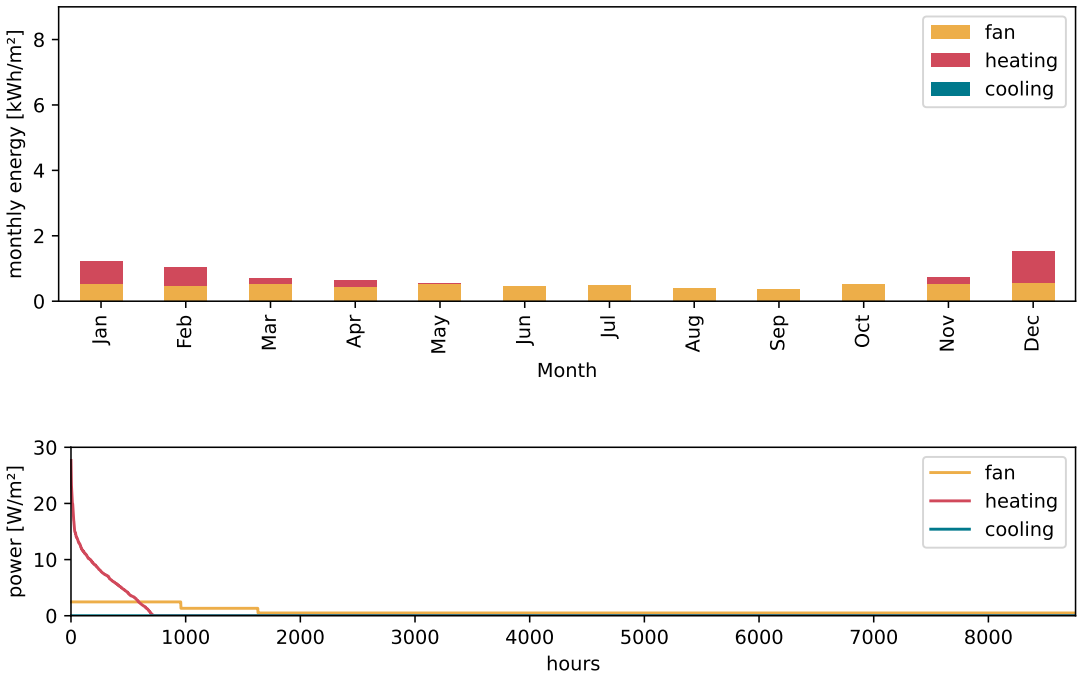


Figure 9.14: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA5

The figure displays the monthly air treatment and transportation energy consumption of the ventilation control strategy IDA5, which is based on occupancy and uses variable air volume. In contrast to control strategy IDA4, the implementation of an occupancy-based control does not result in significant energy reductions. This finding contradicts the results of case study 1 in the Office room, where an occupancy-based control achieved notable reductions compared to a presence-based control. The analysis attributes this difference to the room usage of the Classroom. The Classroom is typically occupied at maximum capacity with 25 pupils and a teacher, and partial room occupancy is uncommon, only occurring during afternoon hours. Consequently, an occupancy-based ventilation control performs similarly to a presence-based ventilation control for most of the time. The annual load duration curve reveals that the ventilation unit operates for 1600 hours per year, with almost 1000 hours at maximum fan rate and 600 hours at a reduced volume flow. The heating system is active for 750 hours, and the peak power demand is 20 W/m^2.

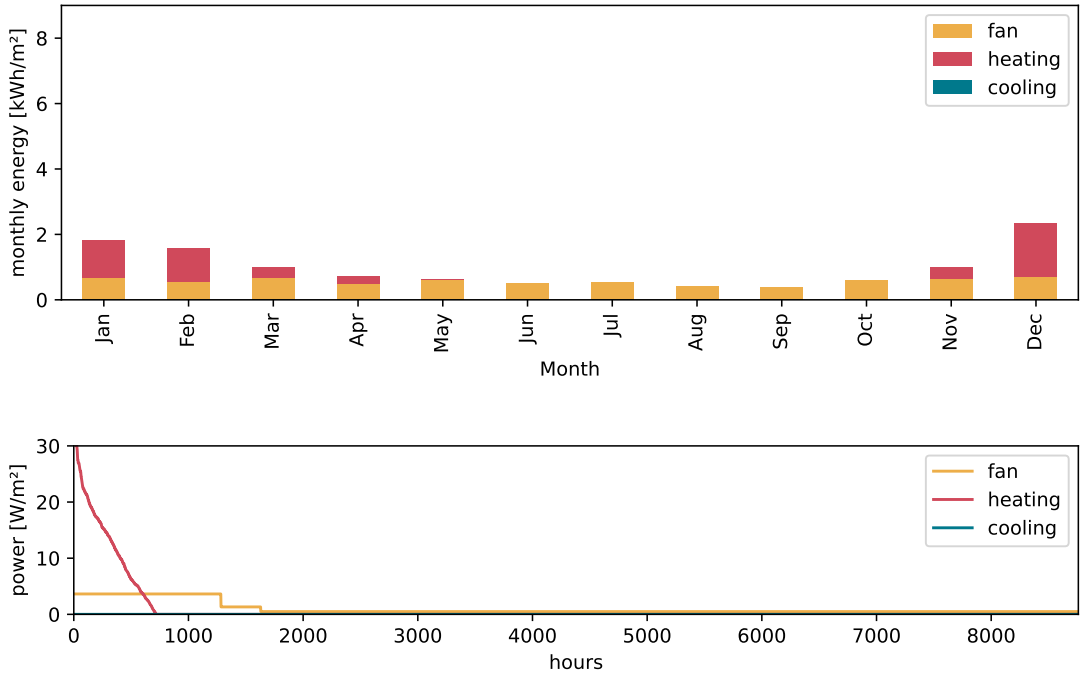


Figure 9.15: Monthly air treatment and transportation energy consumption (top) and annual load duration curve (bottom) in control strategy IDA6

9.6.5 IDA 6

The figure illustrates the monthly air treatment and transportation energy consumption of the ventilation control strategy IDA6, which is based on indoor air pollutant levels. IDA6 incorporates virtual indoor air pollutant sensors. Overall, IDA6 requires 12% more energy than IDA4 and 22% more energy than IDA5. However, it still reduces the energy consumption of IDA1 and IDA3 by 79% and 50% respectively. The increase in energy consumption compared to occupancy and presence-based ventilation control can be attributed to the frequent exceedance of indoor air pollutant thresholds, which necessitate high-volume flows for dispersion. The annual load duration curve demonstrates that the ventilation unit operates for 1700 hours. Of these, 1300 hours are at full load and 400 hours are at part load. Due to the higher volume flow during high pollutant events, the peak heating power increases to 30 W/m² compared to previous ventilation control strategies.

9.6.6 Comparison

The annual air treatment and transportation energy consumption of the decentralized ventilation unit, categorized by heating, cooling, and fan energy, is illustrated in the figure. The implementation of an occupancy-based ventilation control yields the highest

energy reductions, closely followed by presence-based and indoor pollutant-based control strategies. Conversely, IDA1 and IDA3 ventilation control strategies exhibit significantly higher energy consumption. Thus, if the sole consideration is energy consumption, a presence or occupancy-based control strategy offers the greatest reduction potential in the Classroom. However, if the goal is to minimize indoor air pollutant concentrations alongside energy consumption, control strategy IDA6 yields the best results. Additionally, the previous case study demonstrated that ventilation strategy IDA6 performs efficiently in low occupancy scenarios and adapts to different occupancy patterns.

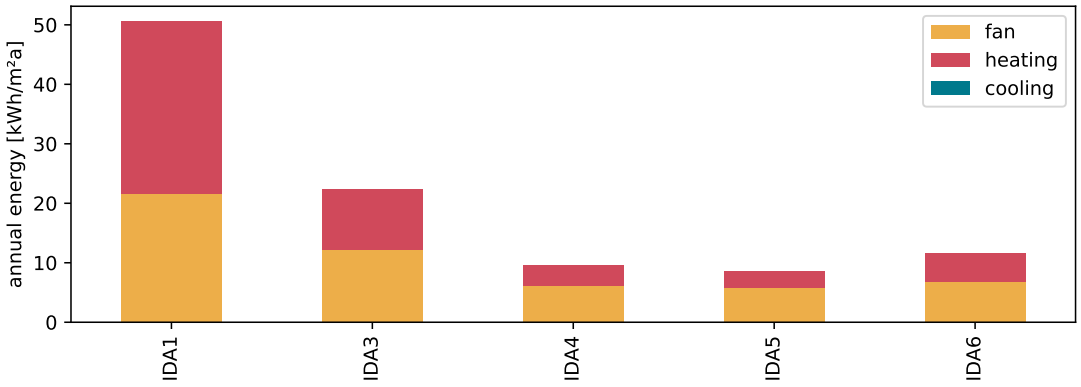


Figure 9.16: Annual air treatment and transportation energy consumption in control strategies IDA1-IDA6

10 Case Study - Comparison and Conclusion

In this final section of Part 4, the results of both case studies are recapped and compared to each other. Figure 10.1 presents the annual air treatment and transportation energy

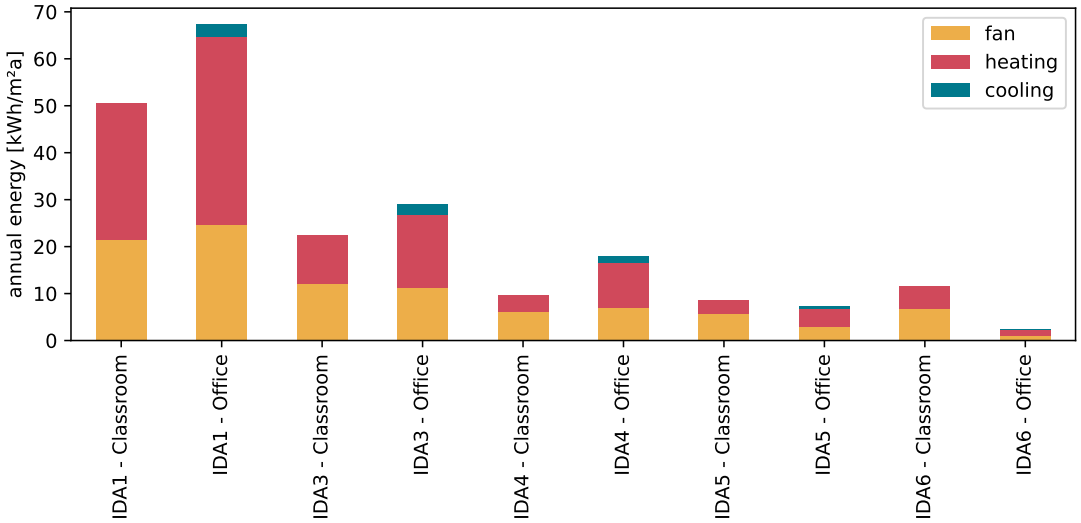


Figure 10.1: Annual energy consumption in control strategies IDA1-IDA6 for rooms office and Classroom

consumption results for each ventilation control strategy (IDA1 - IDA6) and room, specifically in terms of heating, cooling, and ventilation energy consumption. It is evident that the ventilation energy consumption is similar for both rooms in ventilation control strategies IDA1 - IDA4. However, the heating demand is significantly higher for these strategies in the office room due to lower internal gains during occupation and lower people density. With control strategies IDA3 and IDA4, the difference in heating energy demand between the Office and Classroom increases. This is because the internal gains of the Classroom can be utilized more efficiently by operating the ventilation primarily during occupancy hours. For control strategies IDA5 and IDA6, the office room experiences a significant decrease in annual energy consumption and falls below the Classroom. This difference is attributed to the varying occupancy patterns in the office and Classroom. The Classroom is mostly fully occupied during presence with minimal increments, while the office is never fully occupied and experiences incremental variations in occupancy for days and hours. Therefore, occupancy-based control (IDA5) with variable airflow based

on the number of occupants in the room offers a higher reduction potential in the partly occupied office room compared to the fully occupied Classroom. The same principle applies to indoor air pollutant-based control. Since occupancy was identified as one of the main drivers of indoor air pollutants, IDA6 greatly reduces energy consumption in the office room due to the overall low pollution level caused by low occupancy. In contrast, in the Classroom, the IDA6 ventilation control strategy increases energy consumption compared to the presence and occupancy-based control strategies, as the volume flow rates in those strategies were insufficient to remove indoor air pollutants. To summarize, the IDA6 ventilation control strategy, which integrates virtual indoor air pollutant sensors, efficiently reduces air treatment and transportation energy consumption in partly occupied environments such as open office rooms. However, in rooms with high occupancy density like classrooms, an increase in energy consumption was observed compared to occupancy and presence control strategies. This increase can be attributed to the need for increased airflow to effectively remove indoor air pollutants and maintain a healthy indoor air quality. Finally, demand-controlled ventilation strategies IDA4 - IDA6 significantly reduce ventilation runtime by activating the ventilation system only when necessary. This reduction in runtime not only saves air treatment and transportation energy but also extends the device lifetime and maintenance intervals. As a result, costs are reduced by prolonging maintenance intervals, minimizing acquisition costs for consumable supplies like filters, and reducing the workload for maintenance workers. In the case of IDA6, these benefits can be achieved without additional sensors or technical equipment by utilizing virtual indoor air pollutant sensors based on existing BMS data.

Part V

Discussion and Conclusion

11 Discussion

The discussion of this dissertation covers a summary of the main findings, an interpretation of the results, a comparison with related works, an analysis of the strenghts and limitations of this dissertation, implications for practice and policy as well as recommendations for future research.

11.1 Summary of Key Findings

This dissertation explores the development and applicability of multi-pollutant indoor air pollutant virtual sensors in non-residential buildings, with a focus on CO_2, $PM_{2.5}$, and VOC concentrations. Two specific room typologies — open offices and classrooms were examined in detail, constructing a multi-year dataset spanning numerous rooms, buildings, and typologies. The virtual indoor air pollutant sensors were developed using machine learning techniques, specifically Long Short-Term Memory (LSTM) models, and compared to other machine learning algorithms, such as Multi-Layer Perceptron (MLP) and Stochastic Gradient Descent (SGD). In addition, two case studies were conducted, testing the application of the virtual sensing model in demand-controlled ventilation systems. The main findings of this research are as follows: LSTM-based virtual sensing demonstrated a high predictive accuracy for indoor air pollutants within the room it was trained in, given a sufficient volume of training data. The LSTM model's performance during the training process showed sensitivity to the amount of data supplied, aligning with findings in other fields where LSTM models have been applied. Best results were achieved using multi-year, high spatiotemporal resolution data. Ensuring that training data span at least one full year is crucial to mitigate potential bias introduced by selecting incomplete annual periods during which a building's performance may differ significantly. Using a dataset comprising two years of measurements at a 10-second resolution to create a virtual indoor air pollutant sensor resulted in a model with a high degree of accuracy. The mean absolute error rates aligned with the measurement inaccuracies of physical sensors, thereby validating the potential of virtual sensors as alternatives to their physical counterparts. However, it was found that prediction accuracy diminished when applied to other rooms and typologies due to inherent variations in occupancy patterns, localized pollutant sources, and building materials used. This indicates a need for expanded training datasets encompassing a broad range of rooms, buildings, typologies, and geographical regions to better generalize and adapt the models to diverse environments. The practical application of the virtual sensing model was tested in two case studies, in an open office and a Classroom and demonstrated a significant air treatment and transportation energy reduction potential. Compared to scheduled or constant ventilation, reductions of over 95% air treatment and transportation energy was

achieved, especially in the open office room. This potential is particularly notable in office typologies, as occupancy rates have declined in the wake of the COVID-19 pandemic due to a shift towards remote work. Traditional scheduled ventilation control systems do not accommodate dynamic occupancy patterns, which don't follow a set schedule. Therefore, pollutant-based demand-controlled ventilation systems have shown significant potential for air treatment and transportation energy reduction, especially in these typologies. The literature review indicated a high spatiotemporal variation of indoor air pollutants in non-residential buildings. This suggests the need for ubiquitous indoor air pollutant monitoring to accurately assess individual pollutant exposure. The evaluation of the multi-year dataset corroborated these findings, revealing significant variations in pollutant concentrations between different rooms within the same building, despite the HVAC system operating on the same schedule. Surprisingly, even adjacent rooms demonstrated considerable differences in pollutant concentrations. When comparing different machine learning algorithms, LSTMs consistently outperformed MLP and SGD algorithms in terms of predictive accuracy. However, in transfer learning scenarios with limited data available for model adaptation, the simpler SGD model occasionally matched or slightly outperformed the LSTM model, which requires substantial volumes of training data to optimize its performance. Overall, this dissertation emphasizes that the quantity and diversity of training data are pivotal for the development of robust models capable of generalizing effectively across a broad range of environments.

11.2 Interpretation of Results

The results of this dissertation carry practical implications for the application of machine learning methods in virtual indoor air pollutant sensors, particularly for real-time pollutant exposure monitoring and improving building operations. Most notably, the conclusion that virtual indoor air pollutant models can feasibly replace physical sensors provides the potential for scalable, low-cost deployment across a wide range of buildings within a short time frame. Compared to the installation of physical sensors, virtual indoor air pollutant sensors merely require integration with the building management system and outdoor meteorological and pollutant data interfaces. This enables the deployment of indoor air pollutant monitoring in buildings where costs have previously prohibited physical sensor installation. Furthermore, the absence of maintenance and replacement further supports the viability of this approach. The theoretical section of this dissertation explored the health implications of various pollutants and their emergence and spatiotemporal distribution in non-residential buildings' indoor environments. Accurately assessing an individual's pollutant exposure necessitates high-resolution spatiotemporal monitoring, as centralized measurements cannot accurately reflect pollutant concentrations across diverse rooms. However, it was found that physical sensor equipment is not feasible for ubiquitous monitoring due to the careful deployment, continuous maintenance, and frequent replacement needed because of sensor failure, sensor drift, and pollutant accumulation within the sensors. Consequently, virtual sensing of indoor air pollutants presents a viable alternative for personalized pollutant exposure monitoring in indoor

environments. Given that the pollutant data from virtual sensors is available at a high spatiotemporal resolution, a detailed exposure map can be created for each occupant based on their time spent and location within the indoor environment. In practice, this could be implemented via an individual pollutant tracking app that records time and location within a building, provides feedback on individual pollutant exposure, and sends notifications when health-relevant thresholds are exceeded, facilitating early intervention. This approach would enable individuals to access their long-term pollutant exposure data, spanning multiple indoor environments, thus accurately assessing their health risks induced by pollutant exposure. Measurements taken in four office rooms, along with literature data, revealed that office space occupancy patterns underwent significant changes during and after the COVID-19 pandemic. Rates of full and partial work-from-home increased dramatically, leaving many office spaces partially occupied. Yet, HVAC operations could not adapt to these more dynamic occupancy patterns, as operations are often based on static schedules that do not permit demand-based air supply. In the office building examined in this dissertation, none of the four open office rooms ever reached 100% occupancy, leading to significant oversupply of air as the settings were static for the entire building and operated on a fixed schedule. This was evident from the consistently low CO_2 concentration, which rarely exceeded 600 ppm. This suggests a substantial air treatment and transportation energy reduction potential if slightly higher pollutant concentrations were allowed by reducing the supply air volume and controlling the supply air based on pollutant levels rather than a fixed schedule. The potential energy reduction was found to be as high as 96% in a case study for one open office room in this building. These findings are considered transferable to other office buildings, illustrating the energy reduction potential of HVAC systems that could be realized by deploying virtual indoor air pollutant sensors for demand-controlled ventilation. Although many non-residential buildings—especially office buildings—are equipped with HVAC systems, natural ventilation still plays a major role in countries like Germany, where most buildings offer operable windows alongside an HVAC system. Natural ventilation can either be a source or a sink of pollutants, depending on the outdoor air's pollutant content, which dynamically changes and depends on multiple factors. Thus, virtual sensing of indoor air pollutants, accounting for outdoor meteorological and pollution data, can be used to inform occupants and provide recommendations about whether to close or open operable windows to optimize indoor air quality. Especially for buildings situated next to heavily trafficked roads, outdoor air pollution is heavily dependent on the time of day, traffic volume, wind direction, and weather conditions and may introduce harmful pollutants into indoor environments if windows are opened. Conversely, the buildup of pollutants from indoor sources such as furniture and equipment might necessitate opening windows. In practice, this could be implemented with a traffic-light-like indicator system, guiding occupants to open and close windows based on the virtual sensing model.

11.3 Comparison with Related Works

Previous research has recognized the variability of indoor air pollutants in non-residential buildings, presenting diverse approaches to deploying low-cost indoor air pollutant monitors for widespread monitoring. Additionally, some studies have examined the feasibility of model-based assessments of indoor air pollutants, either using a white-box approach by modelling physical pathways of pollutants or a black-box approach employing Stochastic Gradient Descent (SGD) and Multi-Layer Perceptron (MLP) models. However, many studies that inspected indoor air pollutant concentrations in non-residential buildings relied on short measurement periods (ranging from days to weeks) and drew their findings and conclusions from small datasets, thereby introducing significant bias. None of the reviewed studies undertook multi-year measurements of indoor air pollutants; therefore, literature lacks data regarding seasonality or long-term trends. Moreover, most of the examined studies performed measurements at an inadequate sampling rate, neglecting transient responses of indoor air pollutant concentrations. While high spatial resolution was acknowledged, only one study scrutinized the spatial distribution of pollutants, leaving a gap in the literature regarding the spatial distribution of indoor air pollutants in other non-residential building typologies. Additionally, most model-based approaches were restricted to a single pollutant, neglecting multiple pollutants, thus leading to partial results. Finally, existing research has not yet considered the implications of employing these models in demand-controlled ventilation concerning energy consumption. This dissertation addresses these gaps in existing literature by constructing a dataset of indoor air pollutants with long-term, high spatiotemporal resolution data from multiple rooms and building types, surpassing the indoor air pollutant datasets currently used in literature. Using this dataset, an in-depth analysis of the indoor air pollutant status across different rooms was performed and utilized to construct a machine learning model that acts as a virtual indoor air pollutant sensor for $PM_{2.5}$, CO_2, and VOC. A LSTM approach was selected and the evaluation of the virtual sensing model affirmed its capability to replace physical indoor air pollutant sensors, demonstrating very high accuracy in the testing set for the training room, as well as acceptable accuracy when transferred to other rooms of the same building type. Finally, two case studies provided a practical perspective on the potential air treatment and transportation energy savings that could be unlocked by deploying virtual sensing models for indoor air pollutants in demand-controlled ventilation.

11.4 Strengths and Limitations

This section discusses the strengths and limitations of the dissertation.

Strengths The findings of this dissertation are founded on a comprehensive dataset characterized by high spatiotemporal resolution, surpassing other published datasets in the field. The machine learning model was developed using two years of measurement

data sourced from multiple measurement nodes at ten-second intervals. Consequently, nearly 2 billion data points were recorded across all rooms used for training, testing, and transferability testing. To ensure precise measurements at this high rate, custom indoor air pollutant nodes were devised to utilize validated sensing equipment and to control sensor behaviours such as auto-calibration and drift correction, which are not offered by commercial sensors. An in-depth comparison of various machine learning algorithms (LSTM, MLP, and SGD) was performed, and their accuracy evaluated on the testing set of the room they were trained in as well as in transfer learning. Additionally, the availability of server infrastructure and GPU access for model training enabled the execution of multiple model configurations, thereby optimizing the hyperparameter settings. Besides the theoretical development of methods and evaluation of model results, the applicability of the virtual sensing model in demand-controlled ventilation was assessed in two case studies, exploring their potential for air treatment and transportation energy reduction by measuring real decentralized ventilation unit power consumption readings and using a calibrated simulation model for an open office room. This allowed a detailed exploration of various ventilation control strategies as defined in DIN EN 16798-3.

Limitations The deployment of the virtual sensor requires the availability of Building Management System (BMS) data in each room under observation, with a defined standard on measurement interval and spatial resolution. This includes the availability of indoor temperature, indoor humidity, indoor illumination, indoor noise, indoor air pressure, window state, and equipment power consumption data at a temporal resolution of at least 15 minutes and a spatial resolution equivalent to that required for indoor air pollutant monitoring (e.g., for each room). The availability of only a subset of these input parameters does not necessarily render the virtual sensor unavailable but progressively reduces the model's accuracy. Along with BMS data, the deployment of the virtual sensor also necessitates the availability of external data, specifically meteorological and outdoor air pollution data in proximity to the evaluated building. Generally, this data is available in urban areas through public or private meteorological stations, as well as through a network of citizen science-based outdoor pollutant measurements. However, this requirement limits the deployability of the virtual sensing model in rural and less densely settled areas where data points in close proximity may be lacking. Moreover, this research was conducted exclusively in German buildings and under German climatic conditions. Therefore, the German typologies of open offices and classrooms may not be transferable to other regions due to variations in building technology and construction specific to Germany. Additionally, the influence of German climate on the study and model training may also limit the transferability of the model to other regions, since meteorological parameters directly entered the model. As the model was trained with a solely German-based dataset, it introduces a certain level of bias. Hence, the results presented should be taken as proof of concept for the method in other regions, and further implementation using international datasets is encouraged. This dissertation was also restricted to non-residential building types, particularly open office rooms and classrooms.

Therefore, the findings as presented may not be wholly transferable to other typologies in the non-residential type or to other building types such as industrial or residential buildings. The measurement period from June 2021 to May 2023 partially overlapped with the COVID-19 pandemic, which led to lockdowns and work-from-home regulations. Consequently, a subset of the measurement period saw unusually low or no occupancy in the examined rooms, although these periods are small within the overall dataset. However, a noticeable trend throughout the measurement period was a shift in office occupancy due to the increase in employees working from home. Thus, the occupancy in all office rooms seldom reached the maximum and mostly hovered around 50% of maximum occupancy. This trend does not currently seem to be abating and suggests a general change in office usage. In contrast, classrooms were affected by the use of CO_2 alarms signaling an exceedance of the 1100 ppm threshold, which somewhat impacted the typical usage pattern of the occupants and resulted in a higher than normal ventilation rate. Following the deactivation of the alarms, significantly higher CO_2 concentrations were observed in the classrooms, especially in the winter months. Lastly, technological and resource constraints limited the size of the dataset, as the cost of sensor equipment did not allow for monitoring a larger number of rooms and buildings. Additionally, the sensor equipment suitable for long-term high interval measurements meant that VOCs could only be measured qualitatively using Metal Oxide Semiconductor (MOS) sensing, rather than laboratory sampling methods - which are capable of breaking down VOC into its individual components. However, no current technology supports long-term differentiated sampling of VOCs in situ.

11.5 Implications for Practice and Policy

The findings of this dissertation have multiple implications for practical building operation, policy, HVAC control, individual exposure monitoring, and the use of operable windows. The integration of virtual indoor air pollutant sensors into existing Building Management Systems (BMS) adds value by effectively utilizing previously underexploited data. Building operators can leverage this data for improved operation and energy reduction, while also providing real-time indoor air pollutant information to occupants and optimizing natural ventilation, thus unlocking potential air treatment and transportation energy savings and health improvements. To unlock this potential, it is imperative to ensure that BMS systems provide the necessary interfaces and data exchange capabilities to employ virtual indoor air pollutant sensors. This could be facilitated by technical regulations or defined standards. The lack of open systems and interfaces can restrict this development; hence, it is crucial to ensure BMS compatibility and effective communication between different applications and the BMS. Furthermore, the deployment of virtual indoor air pollutant sensors can provide actionable insights from pollutant monitoring in practice. This enables processes for monitoring and responding to critical pollutant concentrations in real time, thus reducing the health risks and potential illnesses of occupants. In practice, this could be similar to "Hitzefrei" (a day off due to excessively hot weather) to reduce pollutant exposure, which consequently encourages building operators to prevent

such high pollutant events. Moreover, virtual indoor air pollutant sensing empowers occupants by transparently and openly communicating pollutant levels. This occupant involvement allows occupants to make data-informed decisions regarding their location and time spent in different indoor environments, as well as decisions regarding natural ventilation and window opening. This necessitates unified strategies for presenting the data to ensure usability and understanding without requiring technical knowledge. The most significant benefits are energy and health improvements gained by integrating virtual indoor air pollutant sensors into ventilation control systems. By providing supply air based on demand rather than fixed schedules, these systems call for policies that promote dynamic and demand-based control strategies in new buildings and renovations, as opposed to static ventilation control. By reducing air treatment and transportation energy consumption, pollutant-based demand-controlled ventilation contributes to the overarching goal of decarbonizing the building stock - one of the largest sectors concerning CO_2 emissions. This reduction in HVAC energy consumption doesn't compromise indoor air quality and also improves the lifespan and maintenance periods of HVAC units due to reduced overall runtime. This positive impact on the sustainability of the building throughout its entire life cycle results in direct and indirect cost savings from reduced HVAC runtime. Further work should detail the integration of virtual indoor air pollutant sensors in a step-by-step guide for incorporation into existing BMS infrastructure, addressing potential challenges during integration and their possible solutions. Finally, a continuous monitoring of the integration should quantify the results by calculating success metrics and providing before-and-after comparisons.

11.6 Recommendations for Future Research

The strengths and limitations of this research were explored in a previous section. Based on these limitations, several recommendations for future research can be proposed. This dissertation identified the need for training machine learning models for virtual sensing of indoor air pollutants using extensive datasets. This is essential to ensure the transferability and deployment of the model across various rooms and environments.
Artificial intelligence is a rapidly advancing field, encompassing a wide range of emerging technologies in machine learning. This offers significant potential for future research to incorporate these developments into virtual indoor air pollutant sensors. Such integration might utilize explainable AI methods to enhance the transparency and understanding of how models operate, shifting from opaque black-box models to more interpretable grey-box models. It could also involve adopting the concept of ensemble learning, increasingly used in large language models, for the virtual sensing of indoor air pollutants. This approach would involve creating specialized models for different pollutants and combining them into an overall comprehensive model.
In this dissertation, a comprehensive dataset was compiled, covering multiple rooms and buildings over a two-year span. However, the results regarding transferability indicate room for improvement, suggesting the need for an even more extensive dataset. Future datasets should cover a wider range of building types and geographic regions and maintain

the spatiotemporal resolution proposed in this dissertation.
Furthermore, hybrid methods that mix rule-based AI with probabilistic approaches or combine white-box simulations with machine learning techniques could address some limitations of purely probabilistic models. These methods could mitigate challenges associated with unlearned building physics and unknown building characteristics by simulating these aspects, reducing the reliance on extensive training data and enhancing model adaptability. Employing a semi-supervised learning approach could further improve the generalizability of models by incorporating a substantial amount of unlabeled data. This strategy might enhance the models' applicability and reduce the need for costly ground truth measurements of indoor air pollutants.
Additionally, further research is encouraged to integrate virtual indoor air pollutant sensors into HVAC control systems. Long-term monitoring and evaluation of energy consumption and indoor air quality in comparison to a standard room should be a primary focus. Future work could also explore the robustness and prediction accuracy of virtual indoor air pollutant sensors under different data quality conditions or with missing data. It would be valuable to reimplement models for various combinations of available input data to assess their performance. Future research should explore the applicability of these methods to other indoor environmental conditions, given the successful application of virtual sensors for indoor air pollutant monitoring in non-residential open offices and classrooms. Adapting these techniques to a broader field of application such as outdoor air pollution, industrial processes, industrial air pollution, and other relevant domains is also recommended.
Finally, future research should also critically examine the role of AI in light of advancements in sensing technologies, which may offer improved methods for monitoring indoor air pollutants in diverse buildings. While current sensing technologies are limited in providing long-term, high-resolution measurements affordably and accurately, the field is advancing rapidly. Developments like Micro-Electro-Mechanical Systems (MEMS) in smartphones have made sensing of temperature, accelerometers, and gyroscopes both widespread and reliable. A parallel improvement in indoor air pollutant measurement technologies could similarly revolutionize monitoring and control in indoor environments, potentially enabling personalized pollutant exposure monitoring and making virtual indoor air pollutant sensors a transitional technology.

12 Conclusion

The conclusion of this dissertation covers a recapitulation of the research question and its main findings, a summary of the contribution to the field, the implications to indoor air quality management as well as a future outlook.

12.1 Recap of the Research Questions and Findings

The following section summarises the research objectives and findings. The research questions established in the introduction were as follows: *1. What is the state of indoor air pollution in classroom and open office zones? What are the seasonal variations, correlations, and distribution patterns in these zones, and how do they compare?* A multiyear dataset of four open office zones and one classroom with close to 2 billion datapoints was gathered and analysed in detail in this dissertation. The results showed significant spatiotemporal variations across rooms, buildings, and typologies calling for high resoluted monitoring. A exploration of correlation factors identified varying determinants, however with occupancy and outdoor pollution dominating in the classroom and HVAC operation in the office zones. Seasonality was found to be espescially pronounced for $PM_{2.5}$ in the office typology and for VOC, CO_2 and $PM_{2.5}$ in the classroom. *2. Can machine learning techniques be used to develop multi-pollutant virtual sensors for predicting indoor air pollutant concentrations? Can these virtual sensors be transferred to different zones and building typologies?* Several machine learning methods were examined, with a specific focus on long short-term memory (LSTM) networks. It was determined that given a large dataset of high-quality data, machine learning models can be developed to achieve high accuracy in virtual sensing of indoor air pollutants. Other machine learning algorithms, such as multilayer perceptrons (MLPs) and stochastic gradient descents (SGDs), performed less effectively when supplied with large amounts of data. However, SGD performance was similar to LSTMs when only a small amount of training data was available. Transferability of the model to other rooms, buildings, and typologies was evaluated both for the unadapted model and using a transfer learning method to adapt the model to new environments during a brief tuning period. The results indicated a general transferability within the same typology even without fine-tuning the model. However, for other typologies, transferability was not successful, and the model achieved only low accuracy. *3. Can virtual sensors be integrated into demand-controlled ventilation systems? What are the potential air treatment and transportation energy consumption reductions and associated benefits of using virtual indoor air pollutant sensors in non-residential buildings?* Two case studies assessed the applicability of the virtual indoor air pollutant sensor in demand-controlled ventilation—one through a calibrated simulation model and the other using a decentralized ventilation unit in a Classroom. The case

studies revealed a significant air treatment and transportation energy reduction potential that can be realized by deploying virtual indoor air pollutant sensors. The results from the case studies indicated a potential reduction in air treatment and transportation energy consumption up to 95%, depending on the building's typology and the previously used control method.

12.2 Contributions to the Field

This dissertation advances the field by addressing identified gaps in the literature. In this dissertation, a robust indoor air pollutant dataset was generated, featuring high spatiotemporal resolution data from multiple rooms of varying typologies, thus exceeding the quality of existing indoor air pollutant datasets in literature. This dataset was employed to investigate the present status of indoor air pollution in the studied rooms and buildings. Additionally, it served as training data for the development of virtual indoor air pollutant sensors. These virtual sensors exhibited a high level of accuracy when compared to their physical counterparts and allowed for ubiquitous indoor air pollutant monitoring, thereby eliminating the cost and time expenses associated with deploying and maintaining physical sensor networks. The impacts of implementing these virtual sensors within a demand-controlled ventilation system were assessed in two case studies, which evaluated potential air treatment and transportation energy consumption reduction. It was found that the integration of virtual indoor air pollutant sensors in ventilation control offers significant energy-saving potential.

12.3 Implications for Indoor Air Quality Management

The introductory and theoretical sections of this dissertation examined the health impacts of indoor air pollutants. Previous studies highlighted the importance of reducing and monitoring indoor air pollutants to mitigate severe health impacts, including diseases as serious as cancer and causes of prevalent mortality. Older buildings partially addressed this issue by being less air-tight, hence providing natural ventilation. However, energy efficiency efforts have exacerbated the situation by creating air-tight building envelopes that heavily rely on mechanical ventilation. The 'sick building syndrome', which is associated with diverse health impacts on occupants and accounts for numerous sick leave days, has been partly attributed to indoor air pollutants. Despite the acknowledged importance of indoor air quality over many years, affordable and low-maintenance technology for ubiquitous monitoring of indoor air pollutants was hitherto non-existent. Therefore, this dissertation explored alternatives to physical sensors.
The hypothesis guiding this dissertation was: *Virtual indoor air pollutant sensors, which utilize machine learning models (LSTM, MLP, SGD), have the potential to accurately forecast indoor air pollutant concentrations ($PM_{2.5}$, CO_2, VOC).* As detailed by the findings related to the research questions, the hypotheses have been substantiated. This was demonstrated by the analysis of the indoor air pollutant dataset, the evaluation of the virtual sensing models, and the case studies. The results of the analysis of the indoor air

pollutant dataset indicated that indoor air pollutants exhibit significant spatiotemporal variations across rooms, buildings, and typologies. The results of the evaluation of the virtual sensing models demonstrated that virtual sensors can replace physical sensors, given the high accuracy achieved in the test set. Furthermore, virtual sensors were found not to suffer from the limitations of physical sensors such as measurement drift or sensor failures, and they can be ubiquitously deployed in buildings with the necessary building management system (BMS) data. The results of the case studies indicated a potential reduction in air treatment and transportation energy consumption of up to 95%, depending on the building's typology and the previously used control method. Significant energy savings are currently particularly relevant, given the dynamic office occupancy patterns emerging in the aftermath of the COVID-19 pandemic, due to a widespread adoption of remote work practices and the demand to decarbonize the building stock.

12.4 Future Outlook

This research underscores the significance of monitoring indoor air pollutants due to their substantial effects on health and well-being. In this dissertation an emphasis was laid on pollutant-concentration and -exposure monitoring in non-residential buildings, with a particular focus on open office rooms and classrooms. This dissertation identifies the limitations of physical indoor air pollutant sensors, explaining why they fall short for ubiquitous indoor air pollutant monitoring. It posits virtual sensing of indoor air pollutants as an alternative and investigates the limitations of widespread pollutant monitoring on health and energy demand, especially within the context of demand-controlled ventilation systems. The findings and case studies corroborate the proposed hypothesis. Nonetheless, certain limitations discussed in preceding sections raise essential research questions for future exploration, particularly concerning the generalizability and transferability of the virtual sensing models. For virtual indoor air pollutant sensors to gain traction in practice, policy recommendations for stakeholders have been compiled in the Discussion section. These pertain to Building Management System (BMS) compatibility and occupant engagement. Anticipated outcomes of widespread application include improved occupant health and substantial reductions in HVAC energy consumption and maintenance needs, owing to diminished operating hours. The implications of this research implore diverse stakeholders to take action. Building owners and operators are encouraged to integrate virtual indoor air pollutant sensors into their BMS systems, occupants to engage with real-time exposure data and respond appropriately, and policy makers to enact regulations promoting the compatibility and usability of BMS data and establishing standardized interfaces. Furthermore, the methods proposed in this research could be applied more broadly, potentially extending to other building typologies, such as residential or industrial settings, or even to domains like outdoor air pollution. Concluding this dissertation with a quote, "We do not seem to recognize that our real customer is the occupant, not the building" (Peter E. Levy), recent efforts to construct airtight buildings and introduce mechanical ventilation systems should not solely focus on the building itself. Rather, they should prioritize the health

and comfort of the occupants. Thus, future developments in virtual indoor air pollutant sensors must facilitate user involvement, accommodate natural ventilation and operable windows, and empower the user through information rather than wresting control away. This approach can align with demand-controlled ventilation systems, which offer an energy-efficient means of supplying necessary ventilation and can complement natural ventilation. Such a hybrid system is particularly essential in locations with substantial outdoor pollution, where reliance on natural ventilation is not always feasible. Consequently, a balance between user involvement and ubiquitous indoor air pollutant monitoring can yield maximal benefits for health, comfort, and energy consumption.

Bibliography

Sabah Ahmed Abdul-Wahab, Stephen Chin Fah En, Ali Elkamel, Lena Ahmadi, and Kaan Yetilmezsoy. A review of standards and guidelines set by international bodies for the parameters of indoor air quality. *Atmospheric Pollution Research*, 6(5):751–767, 2015.

Abdul Afram and Farrokh Janabi-Sharifi. Theory and applications of hvac control systems–a review of model predictive control (mpc). *Building and Environment*, 72: 343–355, 2014.

Zakia Afroz, Gary Higgins, GM Shafiullah, and Tania Urmee. Evaluation of real-life demand-controlled ventilation from the perception of indoor air quality with probable implications. *Energy and Buildings*, 219:110018, 2020.

Kaiser Ahmed, Jarek Kurnitski, and Piia Sormunen. Demand controlled ventilation indoor climate and energy performance in a high performance building with air flow rate controlled chilled beams. *Energy and Buildings*, 109:115–126, 2015.

Jaehyun Ahn, Dongil Shin, Kyuho Kim, and Jihoon Yang. Indoor air quality analysis using deep learning with sensor data. *Sensors*, 17(11):2476, 2017.

Mohamed Alhashme and Nasser Ashgriz. A virtual thermostat for local temperature control. *Energy and buildings*, 126:323–339, 2016.

Maria Justo Alonso, Henrik Madsen, Peng Liu, Rikke Bramming Jørgensen, Thomas Berg Jørgensen, Even Johan Christiansen, Olav Aleksander Myrvang, Diane Bastien, and Hans Martin Mathisen. Evaluation of low-cost formaldehyde sensors calibration. *Building and Environment*, 222:109380, 2022.

Jae-Yoon An, Sumin Kim, Hyun-Joong Kim, and Janghoo Seo. Emission behavior of formaldehyde and tvoc from engineered flooring in under heating and air circulation systems. *Building and Environment*, 45(8):1826–1833, 2010.

Prashant Anand, David Cheong, Chandra Sekhar, Mattheos Santamouris, and Sekhar Kondepudi. Energy saving estimation for plug and lighting load using occupancy analysis. *Renewable Energy*, 143:1143–1161, 2019a.

Prashant Anand, Chandra Sekhar, David Cheong, Mattheos Santamouris, and Sekhar Kondepudi. Occupancy-based zone-level vav system control implications on thermal comfort, ventilation, indoor air quality and building energy efficiency. *Energy and Buildings*, 204:109473, 2019b.

Joshua S Apte, Kyle P Messier, Shahzad Gani, Michael Brauer, Thomas W Kirchstetter, Melissa M Lunden, Julian D Marshall, Christopher J Portier, Roel CH Vermeulen, and Steven P Hamburg. High-resolution air pollution mapping with google street view cars: exploiting big data. *Environmental science & technology*, 51(12):6999–7008, 2017.

Mohammed Arif, Martha Katafygiotou, Ahmed Mazroei, Amit Kaushik, Esam Elsarrag, et al. Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature. *International Journal of Sustainable Built Environment*, 5(1): 1–11, 2016.

Arva Arsiwala, Faris Elghaish, and Mohammed Zoher. Digital twin with machine learning for predictive monitoring of co2 equivalent from existing buildings. *Energy and Buildings*, 284:112851, 2023.

ASHRAE. *ASHRAE handbook fundamentals*. ASHRAE, Atlanta, Ga., si edition edition, 2017. ISBN 9781523113514.

ASHRAE. *ASHRAE handbook Heating Ventilating and Air-Conditioning Systems and Equipment*. ASHRAE, Atlanta, GA, si edition edition, 2020. ISBN 9781523135080.

Thomas Auer, Lukas Lauss, Karl Martin Heissler, Johannes Maderspacher, Dirk Reiß, Jan Mehnert, Bernhard Rumpe, Sebastian Stüber, Matthias Hannen, Stefan Plesser, Claas Pinkernell, Alex Kröker, and Roland Gentemann. Big data in der gebäudeautomation. Technical report, München, 2020.

Marek Badura, Piotr Batog, Anetta Drzeniecka-Osiadacz, and Piotr Modzel. Evaluation of low-cost sensors for ambient pm2. 5 monitoring. *Journal of Sensors*, 2018, 2018.

Birgitta Berglund, Belt Brunekreef, H Knöppe, T Lindvall, Marco Maroni, L Mølhave, and P Skov. Effects of indoor air pollution on human health. *Indoor air*, 2(1):2–25, 1992.

C Borrego, AM Costa, J Ginja, M Amorim, M Coutinho, K Karatzas, Th Sioumis, N Katsifarakis, K Konstantinidis, S De Vito, et al. Assessment of air quality microsensors versus reference methods: The eunetair joint exercise. *Atmospheric Environment*, 147: 246–263, 2016.

Carmen Cacho, G Ventura Silva, Anabela O Martins, Eduardo O Fernandes, Dikaia E Saraga, Chrysanthi Dimitroulopoulou, John G Bartzis, Diana Rembges, Josefa Barrero-Moreno, Dimitrios Kotzias, et al. Air pollutants in office environments and emissions from electronic equipment: A review. *Fresen. Environ. Bull*, 22(9), 2013.

Avril Challoner and Laurence Gill. Indoor/outdoor air pollution relationships in ten commercial buildings: Pm2. 5 and no2. *Building and Environment*, 80:159–173, 2014.

Avril Challoner, Francesco Pilla, and Laurence Gill. Prediction of indoor air exposure from outdoor air quality using an artificial neural network model for inner city commercial buildings. *International journal of environmental research and public health*, 12(12): 15233–15253, 2015.

H Chojer, PTBS Branco, FG Martins, MCM Alvim-Ferraz, and SIV Sousa. Development of low-cost indoor air quality monitoring devices: Recent advancements. *Science of The Total Environment*, page 138385, 2020.

Darius Ciuzas, Tadas Prasauskas, Edvinas Krugly, Ruta Sidaraviciute, Andrius Jurelionis, Lina Seduikyte, Violeta Kauneliene, Aneta Wierzbicka, and Dainius Martuzevicius. Characterization of indoor aerosol temporal variations for the real-time management of indoor air quality. *Atmospheric environment*, 118:107–117, 2015.

James R Coleman and Forrest Meggers. Sensing of indoor air quality—characterization of spatial and temporal pollutant evolution through distributed sensing. *Frontiers in Built Environment*, 4:28, 2018.

Ingrid Demanega, Igor Mujan, Brett C Singer, Aleksandar S Andelkovic, Francesco Babich, and Dusan Licina. Performance assessment of low-cost environmental monitors and single sensors under variable indoor air quality and thermal conditions. *Building and Environment*, 187:107415, 2021.

Hugo Destaillats, Randy L Maddalena, Brett C Singer, Alfred T Hodgson, and Thomas E McKone. Indoor pollutants emitted by office equipment: A review of reported data and information needs. *Atmospheric Environment*, 42(7):1371–1388, 2008.

Djamel Djenouri, Roufaida Laidi, Youcef Djenouri, and Ilangko Balasingham. Machine learning for smart building applications: Review and taxonomy. *ACM Computing Surveys (CSUR)*, 52(2):1–36, 2019.

Maher Elbayoumi, Nor Azam Ramli, and Noor Faizah Fitri Md Yusof. Development and comparison of regression models and feedforward backpropagation neural network models to predict seasonal indoor pm2. 5–10 and pm2. 5 concentrations in naturally ventilated schools. *Atmospheric Pollution Research*, 6(6):1013–1023, 2015.

G Eranna, BC Joshi, DP Runthala, and RP Gupta. Oxide materials for development of integrated gas sensors—a comprehensive review. *Critical Reviews in Solid State and Materials Sciences*, 29(3-4):111–188, 2004.

Louise Bøge Frederickson, Emma Amalie Petersen-Sonn, Yuwei Shen, Ole Hertel, Youwei Hong, Johan Schmidt, and Matthew Stanley Johnson. Low-cost sensors for indoor and outdoor pollution. *Air Pollution Sources, Statistics and Health Effects*, pages 423–453, 2021.

Matthias Fuchs, Thomas Stark, Martin Zeumer, and Manfred Hegger. *Energy manual: sustainable architecture*. DETAIL-Institut für internationale Architektur-Dokumentation GmbH & Co. KG, 2008.

Martin Gabriel and Thomas Auer. Indoor air pollution estimation using machine learning (ann and svr) in smart buildings. In *Proceedings of BauSim Conference 2022: 9th Conference of IBPSA-Germany and Austria*, volume 9 of *BauSim Conference*,

Weimar, Germany, September 2022. IBPSA-Germany and Austria. doi: https://doi.org/10.26868/29761662.2022.24. URL `https://publications.ibpsa.org/conference/paper/?id=bausim2022_Gabriel_Martin`.

Martin Gabriel and Thomas Auer. Lstm deep learning models for virtual sensing of indoor air pollutants: A feasible alternative to physical sensors. *Buildings*, 13(7):1684, 2023.

Gaëlle Guyot, Max H Sherman, and Iain S Walker. Smart ventilation energy and indoor air quality performance in residential buildings: A review. *Energy and Buildings*, 165: 416–430, 2018.

Quang Phuc Ha, Santanu Metia, and Manh Duong Phung. Sensing data fusion for enhanced indoor air quality monitoring. *IEEE Sensors Journal*, 20(8):4430–4441, 2020.

Freja Hasager, Joachim Dithmer Bjerregaard, James Bonomaully, Hasse Knap, Alireza Afshari, and Matthew Stanley Johnson. Indoor air quality: Status and standards. *Air Pollution Sources, Statistics and Health Effects*, pages 135–162, 2021.

Amirhossein Hassani, Núria Castell, Ågot K Watne, and Philipp Schneider. Citizen-operated mobile low-cost sensors for urban pm2. 5 monitoring: field calibration, uncertainty estimation, and application. *Sustainable Cities and Society*, 95:104607, 2023.

Gerhard Hausladen and Karsten Tichelmann. *Interiors construction manual: integrated planning, finishings and fitting-out, technical services*. Walter de Gruyter, 2012.

Leonhard Heindel, Peter Hantschke, and Markus Kästner. A virtual sensing approach for approximating nonlinear dynamical systems using lstm networks. *PAMM*, 21(1): e202100119, 2021.

Christian Hepf, Lennard Overhoff, Sebastian Clark Koth, Martin Gabriel, David Briels, and Thomas Auer. Impact of a weather predictive control strategy for inert building technology on thermal comfort and energy demand. *Buildings*, 13(4):996, 2023.

Yejin Hong, Sungmin Yoon, Yong-Shik Kim, and Hyangin Jang. System-level virtual sensing method in building energy systems using autoencoder: Under the limited sensors and operational datasets. *Applied Energy*, 301:117458, 2021.

PJ Irga and FR Torpy. Indoor air pollutants in occupational buildings in a sub-tropical climate: comparison among ventilation types. *Building and Environment*, 98:190–199, 2016.

Dirk Jacob. *Gebäudebetriebsoptimierung: Verbesserung von Optimierungsmethoden und Optimierung unter unsicheren Randbedingungen*. Karlsruhe Institute of Technology, 2012.

Jihoon Jang, Jinmog Han, and Seung-Bok Leigh. Prediction of heating energy consumption with operation pattern variables for non-residential buildings using lstm networks. *Energy and Buildings*, 255:111647, 2022.

Jinglin Jiang, Tianren Wu, Danielle N Wagner, Philip S Stevens, Heinz J Huber, Antonios Tasoglou, and Brandon E Boor. Investigating how occupancy and ventilation mode influence the dynamics of indoor air pollutants in an office environment. *ASHRAE Transactions*, 126(1), 2020.

Dong Hwa Kang, Dong Hee Choi, Seung Min Lee, Myoung Souk Yeo, and Kwang Woo Kim. Effect of bake-out on reducing voc emissions and concentrations in a residential housing unit with a radiant floor heating system. *Building and Environment*, 45(8): 1816–1825, 2010.

Jungho Kang and Kwang-Il Hwang. A comprehensive real-time indoor air-quality level indicator. *Sustainability*, 8(9):881, 2016.

Irene Karijadi and Shuo-Yan Chou. A hybrid rf-lstm based on ceemdan for improving the accuracy of building energy consumption prediction. *Energy and Buildings*, 259: 111908, 2022.

B Khazaei, A Shiehbeigi, and AR Haji Molla Ali Kani. Modeling indoor air carbon dioxide concentration using artificial neural network. *International journal of environmental science and technology*, 16(2):729–736, 2019.

Jimin Kim, Minjin Kong, Taehoon Hong, Kwangbok Jeong, and Minhyun Lee. The effects of filters for an intelligent air pollutant control system considering natural ventilation and the occupants. *Science of the Total Environment*, 657:410–419, 2019.

Jakub Kolarik, Nadja L Lyng, Rossana Bossi, Thomas Witterseh, Kevin M Smith, and Pawel Wargocki. 3.6 response of commercially available metal oxide semiconductor sensors under air polluting activities typical for residences. *Indoor Air Quality Design and Control in Low-Energy Residential Buildings (EBC Annex 68)*, page 47, 2020.

Sam Kubba. *Handbook of Green Building Design and Construction, 2nd Edition.* Butterworth-Heinemann ; Safari, [Erscheinungsort nicht ermittelbar] ; Boston, MA, 2nd edition edition, 2016. ISBN 9780128104439.

Andrew Kusiak, Mingyang Li, and Haiyang Zheng. Virtual models of indoor-air-quality sensors. *Applied Energy*, 87(6):2087–2094, 2010.

Joel Kuula, Milla Friman, Aku Helin, Jarkko V Niemi, Minna Aurela, Hilkka Timonen, and Sanna Saarikoski. Utilization of scattering and absorption-based particulate matter sensors in the environment impacted by residential wood combustion. *Journal of Aerosol Science*, 150:105671, 2020.

Gunter Lauckner and Jörn Krimmling. Raum-und gebäudeautomation für architekten und ingenieure. *Dresden: Springer Fachmedien Wiesbaden GmbH*, pages 175–226, 2020.

M Leidinger, T Sauerwald, W Reimringer, G Ventura, and AJJoS Schütze. Selective detection of hazardous vocs for indoor air quality applications using a virtual gas sensor array. *Journal of Sensors and Sensor Systems*, 3(2):253–263, 2014.

Han Li, Tianzhen Hong, and Marina Sofos. An inverse approach to solving zone air infiltration rate and people count using indoor environmental sensor data. *Energy and Buildings*, 198:228–242, 2019.

Haorong Li, Daihong Yu, and James E Braun. A review of virtual sensing technology and application in building systems. *Hvac&R Research*, 17(5):619–645, 2011.

Jiayu Li, Haoran Li, Yehan Ma, Yang Wang, Ahmed A Abokifa, Chenyang Lu, and Pratim Biswas. Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network. *Building and Environment*, 127:138–147, 2018.

Petra Liedl, Gerhard Hausladen, and Michael de Saldanha. Building to suit the climate: A handbook, 2011.

Hai-Ying Liu, Philipp Schneider, Rolf Haugen, and Matthias Vogt. Performance assessment of a low-cost pm2. 5 sensor for a near four-month period in oslo, norway. *Atmosphere*, 10(2):41, 2019.

Angela Luengas, Astrid Barona, Cecile Hort, Gorka Gallastegui, Vincent Platel, and Ana Elias. A review of indoor air treatment technologies. *Reviews in Environmental Science and Bio/Technology*, 14(3):499–522, 2015.

Nan Ma, Dorit Aviv, Hongshan Guo, and William W Braham. Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renewable and Sustainable Energy Reviews*, 135:110436, 2021.

Corinne Mandin, Marilena Trantallidi, Andrea Cattaneo, Nuno Canha, Victor G Mihucz, Tamás Szigeti, Rosanna Mabilia, Erica Perreca, Andrea Spinazzè, Serena Fossati, et al. Assessment of indoor air quality in office buildings across europe–the officair study. *Science of the Total Environment*, 579:169–178, 2017.

Abhisek Manikonda, Naděžda Zíková, Philip K Hopke, and Andrea R Ferro. Laboratory assessment of low-cost pm monitors. *Journal of Aerosol Science*, 102:29–40, 2016.

LG Mansson, LA Svennberg, and M Liddament. Technical synthesis report. *A Summary of IEA Annex*, 18, 1997.

Marin B Marinov, Nina Djermanova, Borislav Ganev, Georgi Nikolov, and Emilija Janchevska. Performance evaluation of low-cost carbon dioxide sensors. In *2018 IEEE XXVII International Scientific Conference Electronics-ET*, pages 1–4. IEEE, 2018.

Marco Maroni, Bernd Seifert, and Thomas Lindvall. *Indoor air quality: a comprehensive reference book*. Elsevier, 1995.

MI Mead, OAM Popoola, GB Stewart, Peter Landshoff, M Calleja, M Hayes, JJ Baldovi, MW McLeod, TF Hodgson, J Dicks, et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203, 2013.

Daniel Mendoza, Tabitha M Benney, and Sarah Boll. Long-term analysis of the relationships between indoor and outdoor fine particulate pollution: A case study using research grade sensors. *Science of The Total Environment*, page 145778, 2021.

Bart Merema, Muhannad Delwati, Maarten Sourbron, and Hilde Breesch. Demand controlled ventilation (dcv) in school and office buildings: Lessons learnt from case studies. *Energy and Buildings*, 172:349–360, 2018.

Christoph Molnar. *Interpretable machine learning*. Lulu. com, 2020.

James F Montgomery, Stefan Storey, and Karen Bartlett. Comparison of the indoor air quality in an office operating with natural or mechanical ventilation using short-term intensive pollutant monitoring. *Indoor and Built Environment*, 24(6):777–787, 2015.

Mads Mysen, Sveinung Berntsen, Per Nafstad, and Peter G Schild. Occupancy density and benefits of demand-controlled ventilation in norwegian primary schools. *Energy and Buildings*, 37(12):1234–1240, 2005.

Fergus Nicol and Michael Humphreys. Derivation of the adaptive equations for thermal comfort in free-running buildings in european standard en15251. *Building and environment*, 45(1):11–17, 2010.

Xiaobing Pang, Marvin D Shaw, Stefan Gillot, and Alastair C Lewis. The impacts of water vapour and co-pollutants on the performance of electrochemical gas sensors used for air quality monitoring. *Sensors and Actuators B: Chemical*, 266:674–684, 2018.

Zhihong Pang, Pingfan Hu, Xing Lu, Qingsheng Wang, and Zheng O'Neill. A smart co2-based ventilation control framework to minimize the infection risk of covid-19 in public buildings 2.

Alexandros Pantazaras, Mattheos Santamouris, Siew Eang Lee, and MN Assimakopoulos. A decision tool to balance indoor air quality and energy consumption: A case study. *Energy and Buildings*, 165:246–258, 2018.

Sébastien Pecceu, Samuel Caillou, and Romy Van Gaever. Demand controlled ventilation: relevance of humidity based detection systems for the control of ventilation in the spaces occupied by persons, 2018.

Gen Pei, Donghyun Rim, Stefano Schiavon, and Matthew Vannucci. Effect of sensor position on the performance of co2-based demand controlled ventilation. *Energy and Buildings*, 202:109358, 2019.

Basheer Qolomany, Ala Al-Fuqaha, Driss Benhaddou, and Ajay Gupta. Role of deep lstm neural networks and wi-fi networks in support of occupancy prediction in smart buildings. In *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 50–57. IEEE, 2017.

Basheer Qolomany, Ala Al-Fuqaha, Ajay Gupta, Driss Benhaddou, Safaa Alwajidi, Junaid Qadir, and Alvis C Fong. Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE Access*, 7:90316–90356, 2019.

Stuart J Russell and Peter Norvig. *Artificial intelligence a modern approach*. London, 2010.

Jagriti Saini, Maitreyee Dutta, and Gonçalo Marques. Indoor air quality prediction using optimizers: A comparative study. *Journal of Intelligent & Fuzzy Systems*, (Preprint): 1–17, 2020.

RL Sapra. Using r2 with caution. *Current Medicine Research and Practice*, 4(3):130–134, 2014.

Dikaia Saraga, Thomas Maggos, Eman Sadoun, Eleni Fthenou, Hala Hassan, Vasiliki Tsiouri, Sotirios Karavoltsos, Aikaterini Sakellari, Christos Vasilakos, and Konstantinos Kakosimos. Chemical characterization of indoor and outdoor particulate matter (pm2. 5, pm10) in doha, qatar. *Aerosol and Air Quality Research*, 17(5):1156–1168, 2017.

Ioan Sarbu and Calin Sebarchievici. Aspects of indoor environmental quality assessment in buildings. *Energy and buildings*, 60:410–419, 2013.

Usha Satish, Mark J Mendell, Krishnamurthy Shekhar, Toshifumi Hotchi, Douglas Sullivan, Siegfried Streufert, and William J Fisk. Is co2 an indoor pollutant? direct effects of low-to-moderate co2 concentrations on human decision-making performance. *Environmental health perspectives*, 120(12):1671–1677, 2012.

Leonie Scheuring, Ing Bernhard Weller, and Ing Clemens Felsmann. *Kontrollierte natürliche Lüftung in Büro-und Verwaltungsgebäuden: Ein Beitrag zur Steigerung von Energieeffizienz und Nutzerbehaglichkeit*. PhD thesis, Technische Universität Dresden, 2022.

Lawrence J Schoen. Guidance for building operations during the covid-19 pandemic. *ASHRAE Journal*, 5(3), 2020.

Robert R Scully, Mathias Basner, Jad Nasrini, Chiu-wing Lam, Emanuel Hermosillo, Ruben C Gur, Tyler Moore, David J Alexander, Usha Satish, and Valerie E Ryder. Effects of acute exposures to carbon dioxide on decision making and cognition in astronaut-like subjects. *npj Microgravity*, 5(1):17, 2019.

Stefan Silbernagl and Agamemnon Despopoulos. Taschenatlas der physiologie, 6., korrigierte auflage. *Thieme*, 7:906–913, 2003.

JP Skön, M Johansson, M Raatikainen, K Leiviskä, and M Kolehmainen. Modelling indoor air carbon dioxide (co2) concentration using neural network. *methods*, 14(15): 16, 2012.

Kirk R Smith and Sumi Mehta. The burden of disease from indoor air pollution in developing countries: comparison of estimates. *International journal of hygiene and environmental health*, 206(4-5):279–289, 2003.

Alberto González Ortiz;Cristina Guerreiro; Joana Soares. Air quality in europe — 2020 report. Technical report, European Environment Agency, 2020.

Andrej-Nikolai Spiess and Natalie Neumeyer. An evaluation of r2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a monte carlo approach. *BMC pharmacology*, 10(1):1–11, 2010.

Andrea Spinazzè, Davide Campagnolo, Andrea Cattaneo, Patrizia Urso, Ioannis A Sakellaris, Dikaia E Saraga, Corinne Mandin, Nuno Canha, Rosanna Mabilia, Erica Perreca, et al. Indoor gaseous air pollutants determinants in office buildings—the officair project. *Indoor air*, 30(1):76–87, 2020.

Jeff Stein, Anna Zhou, and Hwakong Cheng. Advanced variable air volume system design guide. *California Energy Commission's Public Interest Energy Research (PIER).) USA*, page 316, 2007.

Zhongwei Sun, Shengwei Wang, and Zhenjun Ma. In-situ implementation and validation of a co2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. *Building and Environment*, 46(1):124–133, 2011.

Jan Sundell. On the history of indoor air quality and health. *Indoor air*, 14(s 7):51–58, 2004.

Kwok Wai Tham. Indoor air quality and its effects on humans—a review of challenges and developments in the last 30 years. *Energy and Buildings*, 130:637–650, 2016.

Akira Tiele, Siavash Esfahani, and James Covington. Design and development of a low-cost, portable monitoring device for indoor environment quality. *Journal of Sensors*, 2018, 2018.

Yannic Toschke, Janet Lusmoeller, Lars Otte, Johann Schmidt, Svenja Meyer, Alexander Tessmer, Christian Brockmann, Milena Ahuis, Emma Hüer, Christian Kirberger, et al. Distributed lora based co2 monitoring network–a standalone open source system for contagion prevention by controlled ventilation. *HardwareX*, 11:e00261, 2022.

Sergio Trilles, Pablo Juan, Somnath Chaudhuri, and Ana Belen Vicente Fortea. Data on co2, temperature and air humidity records in spanish classrooms during the reopening of schools in the covid-19 pandemic. *Data in Brief*, 39:107489, 2021.

Ahmad Umar and Yoon-Bong Hahn. *Metal oxide nanostructures and their applications*, volume 5. American Scientific Publ., 2010.

Bjørn Jenssen Wachenfeldt, Mads Mysen, and Peter G Schild. Air flow rates and energy saving potential in schools with demand-controlled displacement ventilation. *Energy and buildings*, 39(10):1073–1079, 2007.

Junqi Wang, Jingjing Huang, Zhuangbo Feng, Shi-Jie Cao, and Fariborz Haghighat. Occupant-density-detection based energy efficient ventilation system: Prevention of infection transmission. *Energy and Buildings*, 240:110883, 2021.

Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

Qiong Wu, WenJian Cai, Xinli Wang, and Anutosh Chakraborty. Dehumidifier desiccant concentration soft-sensor for a distributed operating liquid desiccant dehumidification system. *Energy and Buildings*, 129:215–226, 2016.

Junjing Yang, Alexandros Pantazaras, Karn Ashokkumar Chaturvedi, Arun Kumar Chandran, Mat Santamouris, Siew Eang Lee, and Kwok Wai Tham. Comparison of different occupancy counting methods for single system-single zone applications. *Energy and Buildings*, 172:221–234, 2018.

Sungmin Yoon. Virtual sensing in intelligent buildings and digitalization. *Automation in Construction*, 143:104578, 2022.

Chen Zhang, Michal Pomianowski, Per Kvols Heiselberg, and Tao Yu. A review of integrated radiant heating/cooling with ventilation systems-thermal comfort and indoor air quality. *Energy and Buildings*, page 110094, 2020.

He Zhang, Ravi Srinivasan, and Vikram Ganesan. Low cost, multi-pollutant sensing system using raspberry pi for indoor air quality monitoring. *Sustainability*, 13(1):370, 2021.

Yinping Zhang, Jinhan Mo, Yuguo Li, Jan Sundell, Pawel Wargocki, Jensen Zhang, John C Little, Richard Corsi, Qihong Deng, Michael HK Leung, et al. Can commonly-used fan-driven air cleaning technologies improve indoor air quality? a literature review. *Atmospheric Environment*, 45(26):4329–4343, 2011.

Yang Zhao, Wim Zeiler, Gert Boxem, and Timi Labeodan. Virtual occupancy sensors for real-time occupancy information in buildings. *Building and Environment*, 93:9–20, 2015.