APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Mahsa Pahlavikhah Varnosfaderani

This Dissertation has been read and approved by the examing committee:

Advisor: Arsalan Heydarian

Advisor:

Committee Member: James H. Lambert

Committee Member: Lisa Colosi Peterson

Committee Member: Brad Campbell

Committee Member: Julianne Quinn

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

December 2024

Integrating Indoor Air Quality and Occupancy Data for Optimized Operation of HVAC Systems

---

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Mahsa Pahlavikhah Varnosfaderani

December 2024

Integrating Indoor Air Quality and Occupancy Data for Optimized Operation of HVAC Systems

Mahsa Pahlavikhah Varnosfaderani

Executive Summary

Heating, ventilation, and air conditioning (HVAC) systems account for approximately 40% of the energy consumed in buildings. However, traditional systems are still unable to dynamically adjust ventilation rates based on real-time indoor air quality (IAQ) metrics and occupancy data. This thesis discusses the need for HVAC control strategies that dynamically manage ventilation systems to enhance IAQ and energy efficiency. Through a series of studies, this research evaluates the limitations of using carbon dioxide (CO_2) as the only indicator of IAQ and explores the inclusion of total volatile organic compounds (TVOC) for a more comprehensive IAQ assessment. A four-month study conducted in various indoor spaces, including conference rooms and open-plan offices, revealed that TVOC levels frequently exceeded recommended limits, even when CO_2 levels were within acceptable ranges. Poor IAQ conditions persisted in conference rooms 71% of the time during occupancy, primarily during social events in open spaces, while single-occupancy offices showed lower rates of poor IAQ. These findings underscore the importance of incorporating TVOC metrics alongside CO_2 to better assess and respond to indoor pollutant levels. To optimize HVAC operation further, this work explored occupancy detection using environmental data, identifying TVOC as a useful indicator of occupant presence. Statistical models, including Support Vector Machines (SVM) and Random Forests, were applied to classify occupancy based on measurements of CO_2 and TVOC. Results indicated that

TVOC, in combination with CO_2, provided accurate occupancy insights, allowing HVAC systems to dynamically adjust ventilation rates in response to real-time occupancy status. Additionally, a CNN Bidirectional LSTM (CBLSTM) model successfully estimated the number of occupants in a space, offering further opportunities to fine-tune HVAC operation based on occupancy levels. The study also developed predictive models to forecast IAQ metrics, comparing linear statistical models with deep learning approaches. By leveraging the forecasting capabilities of BLSTM models, the system achieved a high degree of accuracy, predicting CO_2 and TVOC levels up to 30 minutes in advance. This allows HVAC systems to make preemptive adjustments, ensuring pollutant concentrations remain within healthy ranges. The predictive models demonstrated that deep learning approaches offer substantial improvements in accuracy over traditional models, enhancing the system's ability to optimize IAQ in real time. A dynamic, demand-driven ventilation approach was tested and compared to a conventional schedule-based system, demonstrating energy savings and improved IAQ. Dynamic operation reduced average TVOC and CO_2 concentrations from 206.05 ppb and 544.18 ppm to 128.11 ppb and 496.89 ppm, respectively, while reducing total ventilation rates during unoccupied periods from 154.56 CFM to 125.94 CFM. This decrease led to energy savings without sacrificing air quality, highlighting the advantages of a responsive HVAC system over traditional methods. Additionally, two novel indices were developed to assess the performance loss associated with IAQ and thermal comfort conditions on occupants. The IAQ index, incorporating CO_2 and TVOC levels, and the thermal comfort index, using temperature and humidity data, were evaluated during both scheduled and dynamic HVAC operations. Results showed that dynamic operation reduced these indices, with the thermal comfort index dropping from 0.25 to 0.18 and the IAQ index from 0.53 to 0.12, indicating enhanced occupant comfort and well-being under dynamic ventilation. In summary, this research demonstrates that integrating TVOC, advanced occupancy detection, and predictive models into HVAC management strategies enables a more comprehensive approach to IAQ and energy optimization. Dynamic, demand-driven HVAC systems have the potential to create healthier indoor environments, reduce energy consumption, and support sustainable building practices by adapting ventilation to the specific needs of each environment.

Dedication

To my parents, Mehrnoosh and Hamid,

my husband, Amirmahdi,

and my brother, Masoud,

for their unconditional love, support, and encouragement

Acknowledgments

I would like to extend my deepest gratitude to my advisor, Professor Heydarian, for the unwavering support and encouragement throughout my PhD journey. Thank you for always believing in me, challenging me to push my boundaries, and providing a safe space to learn and grow. You have not only been an incredible mentor but also a genuine friend and a constant source of inspiration. Reflecting on the decision to join your lab, I feel immensely grateful for your guidance and kindness. It has been an honor and a joy to work with you. Thank you for taking a chance on me.

I would also like to express my heartfelt gratitude to my committee members, Professor Lambert, Professor Peterson, Professor Campbell, and Professor Quinn, for their invaluable support throughout this journey. Thank you for the insightful discussions, collaborative spirit, and thought-provoking questions that challenged me to think more deeply and broaden my understanding of my research. Your guidance has been instrumental in helping me explore new perspectives, and I am genuinely grateful for each of you.

Lastly, I would like to extend my heartfelt gratitude to the incredible members of the BRAIn Lab at UVA—Arash, Alan, Xiang, Siavash, Amir, Feng, Beatrice, Hamidreza, Arman, and Farzin. Your support and encouragement have been invaluable to me on this journey. Thank you all for being such inspiring colleagues and friends.

Contents

List of Figures xii

List of Tables xv

1 Introduction 1

1.1 Thesis Statement 1

1.2 Motivation 1

1.3 Thesis objectives 3

1.4 Thesis outline 4

2 A longitudinal observational study to evaluate changes of IAQ levels in a commercial building 7

2.1 Abstract 7

2.2 Introduction and Background 8

2.3 Methodology 15

2.3.1 Testbed and Sensing System Description 15

2.3.2 Data Collection 18

2.3.3 Poor IAQ Definition 19

2.3.4 Data Processing and Analysis 20

2.4 Results 21
2.4.1 Single-occupancy Offices 22
2.4.2 Conference Room 26
2.4.3 Open-space Areas 28
2.5 Conclusions and Discussion 32
2.6 Summary of contributions 35

3 Detecting Occupancy status and level using indoor environmental factors 36

3.1 Abstract 36
3.2 Introduction & background 37
3.3 Methodology and Results 40
3.3.1 Detecting occupancy status 40
3.3.2 Detecting occupancy level 45
3.4 Summary of contributions 50

4 Time series forecasting of the IAQ metrics 51

4.1 Abstract 51
4.2 Introduction & background 52
4.3 Methodology and results 58
4.3.1 Statistical models 58

4.3.2 Deep neural network models 60

4.4 Summary of contributions 63

5 Enhancing building ventilation through demand-driven strategies: balancing indoor air quality and ventilation rate 65

5.1 Abstract 65

5.2 Introduction & background 66

5.3 Methodology 75

5.3.1 Data Collection 76

5.3.2 Data Analysis & Predictive Models 79

5.3.3 Experiment 83

5.4 Results 101

5.5 Conclusions and Discussion 112

5.6 Summary of contributions 113

6 Introducing an IAQ index based on the expected performance loss according to the indoor environmental metrics' levels 114

6.1 Abstract 114

6.2 Introduction & background 115

6.3 Methodology and results 119

6.4 Summary of contributions 124

7 Conclusion **125**

7.1 Thesis overview 125

7.2 Recommendations and future directions 127

7.3 Publications 129

Bibliography **130**

Appendices **148**

Appendix A VOC and CO_2 histograms of single-occupancy of offices A and B **149**

List of Figures

2.1 The ventilation control logic and IAQ situation if CO_2 is considered as the only proxy of IAQ . . . 13

2.2 Different locations of the L-LL . . . 15

2.3 IAQ sensors arrangement in the single occupancy offices and the conference room. Green and red boxes: IAQ sensors, red box: the selected location of the IAQ sensor for the data collection. . . . 17

2.4 Histograms of CO_2 and TVOC concentration in single-occupancy office C 22

2.5 IAQ situation in single-occupancy offices and the conference room during working hours . . . 25

2.6 CO_2 and TVOC time series data during a working day in a single-occupancy office . . . 26

2.7 Histograms of CO_2 and TVOC concentration in the conference room 27

2.8 The average CO_2 and TVOC time series data during a working day for four weeks in the conference room . . . 27

2.9 Poor IAQ situation of the open-space areas during working hours . . 29

2.10 Average CO_2 and TVOC time series data during days of four social events in the arena . . . 30

2.11 Average CO_2 and TVOC time series data during days of two social events with approximate 30 attendees . . . 31

2.12 CO_2 and TVOC distribution over the lab during an event 32

3.1 Left: sensors arrangement in the conference room. Purple cubes represent IAQ sensors, the blue cube represents the occupancy counter sensor, the red box is the selected location of the IAQ sensor for the data collection, and the yellow and blue arrows represent air inlet and outlet vents, respectively. Right: IAQ sensor and the occupancy counter sensor 47

3.2 Overview of the machine learning models developed for occupancy status and occupancy counter prediction 49

4.1 Proposed deep structure for IAQ metrics forecasting; N: sequence length, K: data size, i: time steps of forecasting in the future 61

5.1 Left: sensors arrangement in the conference room. Purple cubes represent IAQ sensors, the blue cube represents the occupancy counter sensor, the red box is the selected location of the IAQ sensor for the data collection, and the yellow and blue arrows represent air inlet and outlet vents, respectively. Right: IAQ sensor and the occupancy counter sensor 76

5.2 Proposed dynamic approach for ventilation system's operation 78

5.3 Predictive models for occupancy and indoor air quality parameters . 80

5.4 Histogram of TVOC levels during dynamic and scheduled operation of the HVAC systems 102

5.5 Histogram of CO_2 levels during dynamic and scheduled operation of the HVAC systems 102

5.6 IAQ parameters' level of sample days with different ventilation system operations 103

5.7 IAQ parameters' level of sample days with different ventilation system operations 104

A.1 Histograms of CO_2 and TVOC concentration in office A 149

A.2 Histograms of CO_2 and TVOC concentration in office B 150

List of Tables

2.1 Characteristics of sensors used 16

2.2 IAQ guidelines and productivity impacts 18

2.3 Percentage of time pollutants exceeded recommended levels and poor IAQ occurs during working hours 23

2.4 Percentage of time pollutants exceeded recommended levels and poor IAQ occurs during Off hours 24

3.1 ML models accuracy in predicting the occupancy binary status in different locations 44

3.2 ML models performance for predicting the binary occupancy status in the conference room using the occupancy counter sensor as the ground truth for the target variable 48

5.1 Characteristics of sensors used 77

5.2 Summary of study phases with their respective durations, periods, usages, and ventilation modes. 80

5.3 ML models' performance for occupancy status prediction 81

5.4 Parameters' situations for ventilation system automation 84

5.5 Model Evaluation Metrics 94

5.6 Mean value of IAQ parameters and ventilation rate during scheduled and dynamic operation of ventilation systems 105

5.7 Overview of input parameter combinations and their impact on mean ventilation rates and IAQ detection during dynamic ventilation operation. The first row serves as the baseline, using all parameters for decision-making. The "Energy" column indicates whether a combination uses less energy (✓) or more energy (✗) compared to the baseline, as reflected in the mean ventilation rate (CFM). The "IAQ" column assesses the ability of each combination to detect conditions where either CO_2 or TVOC exceeds acceptable limits, with ✓indicating effective detection and ✗ indicating failure to detect all exceedance scenarios. . 107

5.8 Overview of input parameters and mean ventilation rates during the occupied period of dynamic ventilation operation. 109

5.9 Overview of input parameters and mean ventilation rates during the unoccupied period of dynamic ventilation operation. 109

5.10 Overview of the ventilation rates frequencies in percentage during different combinations of input parameters 110

Chapter 1

Introduction

1.1 Thesis Statement

This thesis focuses on integrating indoor air quality (IAQ) monitoring and occupancy data to dynamically optimize the operation of HVAC systems. By combining real-time pollutant levels (CO_2 and TVOCs), occupancy information, and predictive modeling, it examines the frequency at which CO_2 and TVOCs exceed their limit range in a commercial building and demonstrates how ventilation operations can be improved to enhance IAQ and energy efficiency by considering both pollutants and occupancy.

1.2 Motivation

According to the US Energy Information Administration (EIA), residential and commercial buildings accounted for about 40% of total US energy consumption in 2020. The main source of energy consumption in the buildings is typically heating, ventilation, and air conditioning (HVAC) systems, accounting for approximately 48% of the total energy consumption in buildings. HVAC systems are responsible for maintaining comfortable and healthy temperatures and indoor air quality (IAQ) levels within buildings. The World Health Organization (WHO) [1] has reported IAQ factors as one of the major contributors that can negatively impact the occupants' well-being, health, and performance. As of this importance, WHO defined a set of office-related symptoms, including difficulty concentrating, skin irritation, and fatigue which workers reported in many office buildings [1] as sick building syndrome (SBS). Recent studies have also highlighted that IAQ can significantly impact our productivity, concentration, comfort, and physical and mental health [2, 3, 4, 5]. As a result, for proper ventilation, we need to consider both energy consumption aspects and health perspectives to make an informed ventilation management decision.

Traditional HVAC systems in buildings do not use any form of automation or feedback control to regulate the temperature and IAQ in a building. As there is no feedback control system to regulate the temperature and airflow, the HVAC systems continue to run when the building is unoccupied or when the IAQ is in good condition. In addition, traditional HVAC systems can be less effective at maintaining decent air quality throughout a building by maintaining the pollutants' levels within an accepted range. Therefore, dynamic changes are needed for HVAC management because they allow for adjustments to be made in response to the constantly changing environmental conditions (regarding the IAQ and occupancy status) within a building. By implementing dynamic changes to HVAC management, building operators can optimize the systems' performance to meet the specific needs of the building occupants and reduce energy consumption. For example, if the building is not occupied during certain hours of the day, the HVAC system can be programmed to reduce the air exchange rate during those times to save energy. Additionally, if the IAQ level is poor due to any reason, such as outdoor air quality conditions or specific events happening indoors, the HVAC system can be programmed to adjust its operations accordingly to maintain a consistent IAQ and avoid unnecessary energy consumption.

With these challenges, there is an increasing demand for advanced HVAC control systems with dynamic adjustment mechanisms that would account for real-time IAQ metrics and occupancy with the goal of improving energy efficiency and indoor environmental quality (IEQ). This thesis targets these research shortcomings by studying how dynamic demand-based control strategies for HVAC systems can improve IEQ (and occupants' health and productivity as the result) without compromising energy efficiency. Toward this goal, I will perform a series of studies to investigate the integration of multiple indoor air pollutants, advanced predictive models for occupancy and IAQ-related metrics forecasting, and the development of new integrated indices based on IAQ factors, thermal comfort, and occupancy data to evaluate the efficiency of our proposed method for dynamic operation of the ventilation systems.

1.3 Thesis objectives

This project's overarching goal is to emphasize the need for dynamic approaches for operating ventilation systems and demonstrate the benefits of these approaches. To achieve this research goal, we formulate the following specific objectives:

1. Evaluating the adequacy of using carbon dioxide (CO_2) as the sole indicator of IAQ for HVAC system management in different spaces with different usages and occupancy patterns.

2. Exploring the use of predictive models for occupancy status and level prediction using IEQ metrics as the features.

3. Exploring the use of statistical time series models and deep neural networks to forecast the future values of IAQ metrics.

4. Design a dynamic HVAC control system that optimizes ventilation rates based on real-time IAQ metrics and occupancy data.

5. Create a comprehensive IEQ index that evaluates the efficacy of the designed dynamic HVAC control system

Overall, the chapters in this dissertation proposal demonstrate the contributions of my research on how we can make more informed decisions regarding ventilation operations for achieving optimal IAQ, which improves the occupants' well-being and productivity while minimizing energy consumption and costs.

1.4 Thesis outline

The remaining chapters address the research goals outlined in the previous section.

Chapter 2 establishes the importance of considering both CO_2 and TVOC levels in HVAC management. Although CO_2 is widely used as an indicator of occupancy and IAQ, TVOC, which are emissions from building materials, cleaning products, and human activities, also play an essential role in the occupants' well-being since they relate to symptoms like headaches, fatigue, and respiratory issues. My longitudinal study evaluates both CO_2 and TVOC levels across various building spaces and periods, demonstrating the need for a more comprehensive IAQ assessment for HVAC control.

Chapter 3 focuses on how occupancy detection can be optimized using low-cost Internet of Things (IoT) sensors, such as CO_2, temperature, humidity, and light sensors. In general, the traditional methods rely on CO_2 as an indicator of occupancy; however, my research investigates whether TVOC levels can offer reliable occupancy information and thus improve model performance. Using statistical models like Support Vector Machines (SVM) and Random Forest, I showed that combining multiple environmental factors enhances the accuracy of occupancy detection, enabling more precise ventilation control. Additionally, by using deep neural networks, I investigated the possibility of predicting the number of occupants in the space by having the IAQ metrics as the features. Having information on occupancy status helped us detect the exposure period of the occupants to the bad IAQ during the scheduled operation of the ventilation systems.

Chapter 4 focuses on using predictive models in the forecasting of IAQ metrics. Indoor air pollutant behaviors are usually sophisticated and nonlinear, limiting the efficiency of traditional linear models such as the ARIMA model in forecasting pollutant levels. By comparing statistical models with deep neural networks (LSTM and CNN-LSTM), this chapter showed that deep learning models can better capture these complex relationships, therefore improving the accuracy of forecasts for pollutants like CO_2 and TVOC. These models enable proactive HVAC management, where the system adjusts based on predicted pollutant levels, further optimizing energy consumption and maintaining a healthy indoor environment.

Chapter 5 investigates the application of dynamic, demand-driven ventilation strategies that adjust based on real-time indoor air quality (IAQ) data, specifically focusing on TVOC and CO_2 levels and occupancy patterns. Through a four-month field experiment, this chapter compares the performance of a scheduled-based and dynamic operation of the ventilation systems regarding the indoor pollutants' levels and energy consumption as a function of ventilation rates. The dynamic operation mode integrates monitored IAQ and occupancy data into a building's HVAC system to automate ventilation in real-world conditions. This approach overcomes the limitations of traditional ventilation strategies, which rely primarily on fixed schedules, and instead introduces a dynamic adjustment method to account for multiple pollutants' levels and occupancy status. This experimental analysis revealed that dynamic control reduces both pollutants' concentrations and energy consumption.

Finally, chapter 6 proposes a new performance loss index that incorporates multiple environmental factors, including CO_2, TVOC, PM2.5, temperature, and humidity level with the number of occupants. For this goal, individual indices are defined for each factor by utilizing multiple reference values that are selected based on their specific impact on occupant well-being. Then, by using a geometric mean for index aggregation, the final productivity loss index is defined. This index is then used to compare the efficiency of the dynamic and the scheduled-base operation of the ventilation system by using the data collected in the previous chapter.

Chapter 2

A longitudinal observational study to evaluate changes of IAQ levels in a commercial building

2.1 Abstract

People spend approximately 90% of their time indoors, making effective indoor air quality (IAQ) monitoring crucial for occupants' well-being. Traditional IAQ monitoring primarily focuses on carbon dioxide (CO_2) levels to regulate Heating, Ventilation, and Air Conditioning (HVAC) systems. However, HVAC systems often overlook other critical IAQ metrics, such as total volatile organic compounds (TVOC), which may correspond better to occupant activities in some cases. This naturalistic study, conducted over four months at the University of Virginia, addresses this significant gap by observing changes in TVOC and CO_2 levels across various times, events, and spaces, including conference rooms, single occupancy offices, and common open-space areas. We aimed to determine whether CO_2 can be the only representative of IAQ for dynamically adjusting the ventilation rates within this testbed. A key focus was on poor IAQ instances where CO_2 levels were below the recommended levels, but TVOC concentrations exceeded them, potentially impacting occupants' health and

well-being. Our results revealed that in the studied conference room, poor IAQ conditions prevailed 71% of the time during occupancy, in contrast to lower rates in single occupancy offices (11%, 7%, and 16%). Notably, while social events influenced CO_2 levels less, TVOC levels significantly increased in all open-space areas. These findings challenge the conventional reliance on CO_2 monitoring for IAQ management, highlighting the necessity of incorporating comprehensive IAQ metrics in HVAC systems. The study underscores the critical need for dynamic HVAC systems that adapt to real-time IAQ conditions, a vital step towards enhancing indoor environmental quality in various settings.

2.2 Introduction and Background

On average, Americans spend up to 90% of their time in indoor environments [6]. However, most buildings are not designed or operated to enhance our health and productivity[7]. Among the various factors affecting indoor environmental quality (IEQ), which includes noise, light, and indoor air quality (IAQ), the World Health Organization (WHO) [1] identifies IAQ as a critical element with substantial influence on occupant well-being, health, and performance. Recognizing its significance, WHO has outlined a range of symptoms known as Sick Building Syndrome (SBS), which include difficulty concentrating, skin irritation, and fatigue, commonly reported by occupants of many indoor spaces [8]. Recent research underscores the profound impact of IAQ on productivity, concentration, comfort, as well as physical and mental health [2, 3, 4, 5]. Carbon dioxide (CO_2) and total volatile organic compounds (TVOC) [9, 10] are among the primary indoor air pollutants associated with SBS symptoms such as headache, fatigue [11, 12] and asthma [13]. Other contributing factors include particulate matters such as $PM_{2.5}$ and PM_{10}, which may be more prevalent in residential buildings [14, 15] during activities like cooking, in structures near high-traffic areas [16, 17], or during extreme events such as wildfires [18].

According to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 62.1, the primary source of indoor CO_2 emissions is human metabolism [19]. This gas is frequently used as the sole indicator of IAQ and guides the operation of advanced Heating, Ventilation, and Air Conditioning (HVAC) systems. Numerous studies have confirmed that elevated levels of indoor CO_2 are linked to symptoms such as fatigue and dizziness [20], as well as increases in blood pressure [21]. Furthermore, prior research has established a correlation between high indoor CO_2 levels and occupants' discomfort as well as a decrease in the performance of office workers [22].

On the other hand, TVOC are compounds with high vapor pressure and low water solubility and are emitted as gases from certain solids or liquids. TVOC encompass a range of chemicals, some of which can have both short-term and long-term negative health impacts. Indoor concentrations of many TVOC are consistently higher (up to ten times higher) than outdoors due to emissions from a variety of indoor sources, including paints, cleaning supplies, pesticides, building materials, and office equipment such as printers and copiers [23]. Previous studies have shown that the presence and activities of humans and beauty products could impact TVOC levels in indoor environments [24, 25]. These studies found associations between high levels of TVOC and SBS [26], especially when occupants are feeling negative emotional states such as stress [9]. The link between indoor TVOC levels and occupant well-being has prompted further research aimed at developing reliable testing methods for pollutants emitted by both occupants and the products they use. One notable study by Bartzis et al. [27] evaluated the emission rates of consumer products in indoor settings. The authors selected items from different product classes, such as cleaning agents, air fresheners, and personal care products that are known to emit TVOC, and implemented an inter-laboratory comparison by testing the same items in three different test chambers. Their findings revealed that the studied consumer products can significantly elevate TVOC levels beyond healthy limits. Consequently, the study emphasized the need for standardized emission estimation protocols and called for more extensive, naturalistic data collection concerning human-emitted pollutants in indoor environments.

Several research studies have investigated the changes in IAQ factors across different building types, including schools [28], museums [29, 30], cinemas [31], residential buildings [32, 33, 34], and hospitals [35]. Educational institutes are also considered one of the most important indoor environments to be studied regarding IAQ, given that, after residential settings, students and instructors spend considerable time there. [36, 4, 12, 37]. These studies found that poor IAQ exacerbates allergic diseases and asthma across occupants and decreases their performance given the levels and duration of exposure to poor IAQ [28]. They also identified that indoor air pollutants are influenced by seasonal changes, time of day, and variations in occupancy level and behaviors [38]. For instance, in residential settings, certain activities like cooking and painting have been identified as triggers for elevated levels of specific pollutants, notably TVOC [39].

The level of CO_2 in indoor environments has been studied for various applications, including its relevance to HVAC system performance and its impact on occupant well-being. Asif et al. [40] have evaluated the CO_2 level along with indoor temperature and relative humidity in an academic building over a period from March to June.

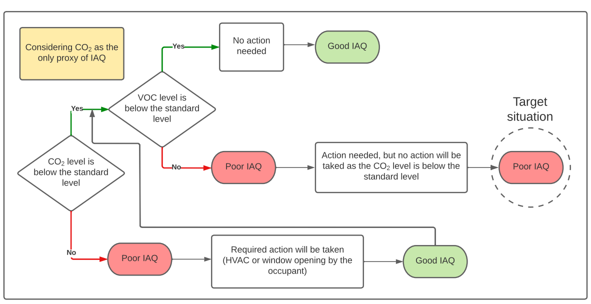
Their research encompassed buildings with diverse HVAC systems and revealed that non-centralized HVAC systems often led to CO_2 levels that exceeded ASHRAE standards during occupied periods. In this study, they only considered CO_2 as the indoor air pollutant and evaluated classrooms with high (or medium) occupancy levels. In another study, Pantelic et al. [41] investigated the metabolically generated CO_2 effects on the "personal cloud" surrounding an individual. This study evaluated the relationship between the metabolic CO_2 concentration and the scenarios typically encountered in the office environment, such as working on a computer and talking. This research showed that occupants' presence and activities result in a higher metabolic CO_2 cloud around the occupants.

Other studies evaluated the relation between CO_2 level and occupants' performance. Several studies showed that exposure to moderate CO_2 levels (1000-2500 ppm) could impair certain attributes of decision making, such as gathering and utilization of information, even in the absence of other bio-effluents [42, 43]. This is while some other studies found exposure to CO_2 at levels up to 3000 ppm without the presence of other indoor air pollutants, such as sources of TVOC, did not cause occupants' cognitive performance degradation. With the presence of other bio-effluents and CO_2 at levels up to 3000 ppm, participants identified acute symptoms such as headache, fatigue, and difficulty in thinking clearly, and decreased certain indicators of cognitive functioning [44]. These findings show the importance of evaluating CO_2 levels alongside other indoor air pollutants, such as TVOC.

A review study by Weschler et al. [45] showed that humans have a pronounced influence on the indoor chemistry they reside in, such as transferring skin oils to indoor surfaces, and specifically highlighted the occupants' influence on ozone levels is significantly large. The study emphasized the need for further research into the impact of human TVOC emissions on the IAQ of enclosed spaces [45]. Another study [46] reviewed methodologies and approaches in previous studies for evaluating TVOC levels. The paper highlighted that to obtain more reliable estimates of human exposure to TVOC, there is a need for naturalistic studies over a long period of time, which can be done using suitable analytical techniques and low-cost sensors. In another study by Wang et al. [47], they implemented a controlled experiment in a climate-controlled chamber occupied by four seated human volunteers. They evaluated the effect of occupants' activities, clothing, age, and relative humidity and temperature on the TVOC's emission rate. Their results showed that the occupants' clothing, as well as instructed movements, such as standing up and stretching, increased the TVOC emission rates. The study also concluded that an increase in temperature and relative humidity results in higher TVOC emissions.

Lin et al. [48] conducted an analysis of the influence of occupant behavior on IAQ, utilizing data from smart home sensors across two testbeds. The study focused on three representative pollutants, $PM_{2.5}$, formaldehyde, and methanol, representing outdoor and indoor pollutants that are associated with indoor materials and occupant activities. This study showed that activities that impact the ambient temperature, such as bathing and cooking, have more impact on the pollutants. This might be because the temperature caused by an activity may last longer than the activity itself; therefore, it can impact the IAQ even after the activity has ended. Other than the effect of occupants on the TVOC level, Allen et al. [43] evaluated the effect of different levels of TVOC on occupants' cognitive performance. In this study, results showed that TVOC and CO_2 are independently associated with the cognitive functions of the studied subjects, and their cognitive functions scores were significantly higher where TVOC and CO_2 levels were lower.

Previous studies have evaluated IAQ in a limited number of spaces for short periods of time. Limited studies capture the changes in IAQ in open space areas in buildings or enclosed spaces with a low number of occupants, such as single-occupancy offices. Additionally, most existing research only focuses on IAQ during occupied periods, neglecting to analyze how metrics fluctuate between occupied and unoccupied times [49]. When studies do account for occupancy, they typically rely on preset schedules for spaces like classrooms to determine occupied hours [40]. Also, limited studies have evaluated the changes in TVOC and CO_2 exposure rates longitudinally. Addressing these gaps by simultaneously evaluating both CO_2 and TVOC levels using affordable IAQ sensors could be a crucial step toward enhancing IAQ and fine-tuning ventilation systems. Over recent years, advancements in accurate and cost-effective Internet of Things (IoT) devices have effectively overcome the scarcity of monitoring indoor air pollutant concentration. Accordingly, by leveraging a network of affordable IAQ sensing systems, we have conducted a longitudinal study of CO_2 and TVOC variations in multiple indoor spaces with diverse functionalities and occupancy profiles.



(a) The flowchart showing the IAQ conditions and the corresponding ventilation actions

TVOC level (200 ppb)	CO_2 level (1000 ppm)	Ventilation activated?	Poor IAQ?
Safe	Safe	No	No
Safe	Poor	Yes	No
Poor	Safe	No	Yes
Poor	Poor	Yes	No

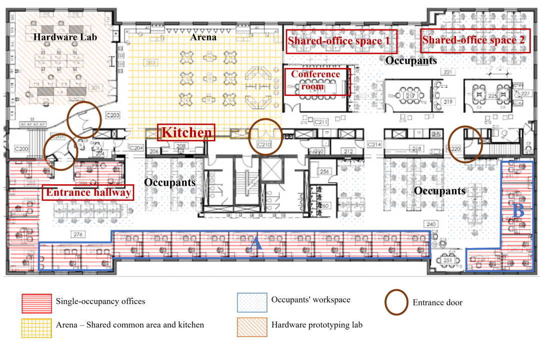
(b) The truth table showing the IAQ conditions based on TVOC and CO_2 levels

Figure 2.1: The ventilation control logic and IAQ situation if CO_2 is considered as the only proxy of IAQ

The objective of our longitudinal study is to investigate whether CO_2 measurements can be the only representative of the IAQ for the purpose of automating ventilation systems or if it's necessary to also consider the levels of TVOC in order to optimize the performance of HVAC systems. Another objective is to evaluate the occupants' approximate exposure frequency to the "poor IAQ" condition in single-occupancy offices and a conference room. Fig. 2.1 shows the rule-based logic for the identification of poor IAQ conditions. As this flowchart shows, poor IAQ conditions could be associated with CO_2 and TVOC concentrations and the consequent ventilation actions. Toward these goals, we gathered data on TVOC and CO_2 levels across various indoor spaces, each characterized by different occupancy patterns. The study areas included three single-occupancy offices, two shared office spaces, a hallway, a kitchen and dining area, and a conference room. Our initial analysis examined the effectiveness of existing ventilation systems in mitigating TVOC and CO_2 concentrations, quantifying the instances where these pollutants surpassed recommended levels. Subsequently, we focused on conditions of "poor IAQ," defined as situations where CO_2 levels remain within the acceptable range, but TVOC levels do not. As noted, Fig. 2.1 shows the occurrence of the poor IAQ situation based on the TVOC and CO_2 levels. We reported the frequency with which the TVOC and CO_2 exceeded their standard level and the occurrence of poor IAQ for the weekdays and weekend schedules of the HVAC system. To augment our analysis, we employed a support vector machine (SVM) model that uses TVOC and CO_2 data to predict occupant presence in single-occupancy offices and the conference room. This predictive model enabled us to quantify the occupants' exposure time to poor IAQ conditions.

2.3 Methodology

This study aims to evaluate whether considering CO_2 as the only proxy of IAQ to inform HVAC ventilation rates is enough for healthier building and indoor environments. This section provides an overview of the testbed (L-LL) and sensing system, data collection, and poor IAQ definition, followed by data processing and analysis.



(a) Different locations of the hallways and the conference room in the lab; single-occupancy offices A and C are selected from zone A, and single-occupancy office B is selected from zone B.

(b) Top left: single-occupancy offices, top right: Arena & the kitchen area, bottom left: conference room, bottom right: shared office spaces

Figure 2.2: Different locations of the L-LL

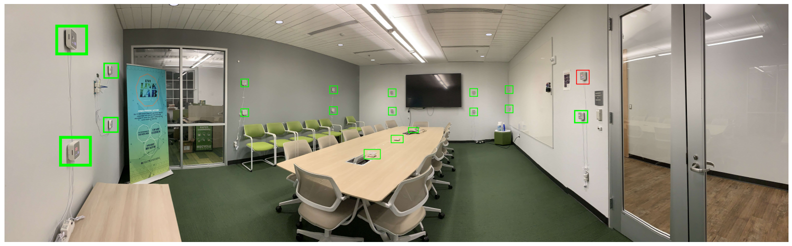
2.3.1 Testbed and Sensing System Description

The data for this study was collected in the L-LL located at the University of Virginia campus. The L-LL is a 17,000 sqft open space office building and is occupied by roughly 150 residents, including faculty, research scientists, graduate students, and staff. L-LL consists of 24 single-occupancy offices, four conference rooms with different capacities, one kitchen and arena space, and four connected main shared office spaces. Fig. 2.2a shows different areas in L-LL, and Fig. 2.2b shows photos of different locations of the testbed, including the arena and the kitchen area, the conference room, single-occupancy offices, and the residents' shared offices. If there are events in this space, they typically take place in the arena, and food and drinks are usually served during events, and occupancy levels can increase by anywhere from 60 to 80 people. These events are typically scientific presentations or social events and gatherings for different celebrations during the year. Since 2018, the space has turned into a living lab where over 350 IoT devices have been deployed to monitor indoor environmental changes, energy consumption, and occupants' interactions with different building systems. These sensors include appliance and electric circuit energy monitors, motion, and indoor environmental quality sensors, such as light, temperature, humidity, noise, CO_2, $PM_{2.5}$, and TVOC levels [50]. Features and parameters of the sensors that are utilized in the L-LL for this study are summarized in Table 2.1. For this study, we used eight Awair Omni IEQ sensors in L-LL, of which three were located in the single-occupancy offices, one in the conference room, one in the hallway, one in each shared office space, and one in the kitchen area to collect the pollutants' concentrations sampled every 10 seconds. In one of the single-occupancy offices and the conference room, one of each type of motion sensor was used to collect the occupancy status for four months (the entire study).

Sensor	Measurement	Value range	Error	Unit
EnOcean E9T-OSW	Motion	[0,255]	-	-
Dual Tech Ceiling Mount Sensor MOS-DT	Motion	[0,255]	-	-
Awair Element	CO_2 TVOC	[400, 5000] [0, 60000]	±75ppm ±10%	ppm ppb

Table 2.1: Characteristics of sensors used

As shown in Fig. 2.3, to optimize sensor placement in single occupancy offices, we initially considered two locations: one closer to the occupant at a seating height of 3.94



(a) Conference room

(b) Single-occupancy office

(c) IAQ sensors

Figure 2.3: IAQ sensors arrangement in the single occupancy offices and the conference room. Green and red boxes: IAQ sensors, red box: the selected location of the IAQ sensor for the data collection.

ft and another further away at a standing height of 3.91 ft. Our analysis indicated that the data from both positions exhibited similar patterns, with readings typically within a 3-5% range of each other. Notably, the sensor positioned closer to the occupant consistently registered slightly higher pollutant levels. Based on these findings, we decided to conduct the experiment using a single sensor placed near the occupant at the 3.94 ft seating height in each room. This positioning provides a more accurate representation of the pollutant concentration levels to which occupants are exposed. To optimize sensor placement in the conference room, we initially installed nineteen sensors at various locations and heights (3.94 and 5.91 ft.) within the room. This pilot experiment was run during working hours (2.6 ACH) for 10 days (separate from the main study). Upon analyzing the collected data, we observed that all sensor readings were remarkably consistent, deviating by only 3-5% from each other. Utilizing both correlation analysis and random forest feature selection techniques, we identified one sensor (shown in red boxes in Fig. 2.3) as the most representative of the collective locations. This particular sensor's significance was further underscored by its strategic positioning, away from the direct influence of any ventilation vents, ensuring that its readings were not artificially impacted. Consequently, this location was selected for the experiment, balancing both statistical rigor and practical considerations in environmental monitoring.

2.3.2 Data Collection

We recorded four months of TVOC, CO_2, and motion sensor data during weekdays and weekends. We divided the data into two different conditions based on the HVAC systems' operation. The first one is the "working hours" condition that includes weekdays from 6:00 AM to 7:00 PM, in which the HVAC systems are preset to operate at a ventilation rate of 2.6 air changes per hour (ACH) across all zones. The second condition is "off hours", which includes the weekends and specific hours of weekdays (7:00 PM- 6:00 AM) when the HVAC operations are preset at a lower air exchange rate of 1.5 ACH due to lower expected occupancy levels.

Particles	Baseline	Moderate	High	Productivity Effect
CO_2 (ppm)	600 [43]	1000 [43, 42]	2500 [42]	-21% for every 400 past 600 [43], -44-94% at 2500 [42]
TVOC (ppb)	50 [43, 51]	200 [52]	500 [43]	-13% at 100 [43]

Table 2.2: IAQ guidelines and productivity impacts

2.3.3 Poor IAQ Definition

We have considered the limit level for both pollutants (CO_2 and TVOC) based on the literature [43, 42, 53, 54, 52] on performance loss of the occupants at different levels of indoor air pollutants. Table 2.2 represents the impact of moderate and high levels of CO_2 and TVOC on productivity loss according to the literature. For this study, we have considered the moderate level category, which considers adverse health and performance impacts at 1000 ppm for CO_2 and 200 ppb for TVOC levels. In this study, we are specifically interested in incidents where CO_2 concentration is below the recommended levels while TVOC levels exceed those recommendations. In these conditions, although the IAQ is poor for the occupants due to high TVOC concentration, the HVAC system is not informed to operate at a higher rate since it is only actuated based on CO_2 levels.

To better understand the ”poor IAQ” condition, Fig. 2.1 represents the potential scenarios that can happen if the ventilation system operates dynamically based on the CO_2 levels. As depicted, poor IAQ can happen in two ways: first, when CO_2 levels surpass recommended thresholds, prompting activation of the HVAC system. Second, when TVOC levels alone exceed the recommended range, while CO_2 remains within acceptable limits, leading to an unresponsive HVAC system. In this paper, the term ”poor IAQ” specifically denotes the latter case. To assess the necessity of incorporating TVOC-level information into HVAC system operation, we quantify occurrences of the aforementioned poor IAQ conditions across various spaces within the L-LL. In essence, this longitudinal study aims to determine whether CO_2 measurements alone suffice for dynamically adjusting ventilation rates, or if the inclusion of TVOC levels is essential for optimizing HVAC system performance.

2.3.4 Data Processing and Analysis

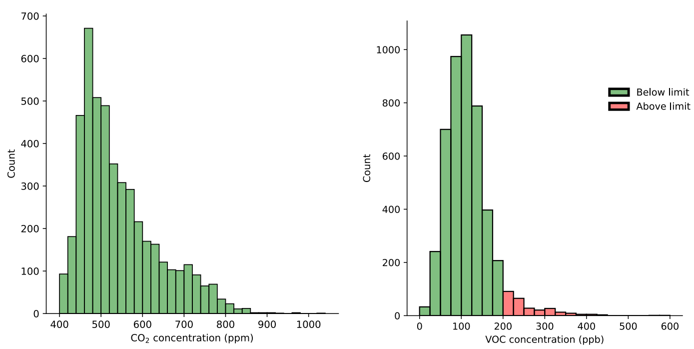
To analyze the data, we aggregated the high-frequency data over 15-minute intervals by computing the average values of the 10-second readings. To effectively identify and remove outliers from the dataset, we replaced readings that fell outside the sensor's operational range (as detailed in Table 2.1) with the mean of the adjacent values, preceding and following the outlier. This approach ensured that any anomalous data potentially resulting from sensor errors did not skew our results. Additionally, we meticulously evaluated the scatterplots of the pollutant time-series data to scrutinize any significantly distant data points from the main cluster. Our thorough examination revealed no such anomalies in the dataset. Also, we installed the sensors one week prior to the main study, ensuring they were adequately calibrated to the environmental conditions of the space.

Additionally, we addressed the missing values by substituting them with the mean of their immediate preceding and subsequent values. We then separated the aggregated data into two categories: working hours and off-hours, identifying periods during which CO_2 and TVOC levels exceeded their respective recommended limits. Additionally, we assessed instances of poor IAQ, as defined in Section 2.3.3, across various locations for both working and off-hour conditions. Extending this analysis to both the kitchen area and the arena, we pinpointed time periods corresponding to scheduled events by referencing the lab's calendar. By analyzing CO_2 and TVOC concentrations during these specific events, we aim to ascertain whether CO_2 alone serves as a sufficient proxy for IAQ under these particular conditions.

As noted, the second objective of our study is to evaluate the occupants' approximate exposure frequency to the "poor IAQ" condition in single-occupancy offices and the conference room. Toward this goal, we trained several machine learning models to predict the binary status of the occupancy levels (0 if there is not anyone in the area, and 1 if there is at least one occupant in the area) by using CO_2 and TVOC levels as the features. We used the motion sensors installed in one of the single-occupancy offices and the conference room for the target variable ground truth. To find the best-fitted model, we trained five different models: SVM, Gaussian naive Bayes (GNB), logistic regression (LGR), random forest classifier (RFC), and k-nearest neighbors (k-NN). The best-fitted model for our data was an SVM model with a non-linear kernel (with an F1 score of 0.86). We evaluated the results using both linear and non-linear kernels for the SVM model, and the one with the Gaussian kernel (as a non-linear kernel) leads to higher accuracy. This model uses the radial basis function (RBF) as the Gaussian kernel. The details of the models we used for the occupancy detection task can be found in our paper [55]. Using the occupancy predictor model helped us find out the percentages of time that occupants were present and exposed to poor IAQ levels in the other two single-occupied offices, which were not equipped with motion sensors. For fine-tuning the SVM model in these offices, we used the occupants' calendar data for a limited time as the occupancy ground truth.

2.4 Results

This section provides the findings from the analysis of the data collected from three single-occupancy offices, a conference room, and shared office spaces, examined both during working hours and off-hours. Subsequent subsections will detail the frequency with which CO_2 and TVOC levels exceed recommended thresholds, as well as instances of poor IAQ – i.e. when TVOC concentrations surpass the recommended level, but CO_2 levels remain within the acceptable range. These subsections show that poor IAQ happens most often in enclosed, shared areas with a high number of occupants, such as conference rooms. Furthermore, our findings highlight that the HVAC system may fail to address the poor IAQ conditions 6 to 30 percent of the time across different zones and space types. These findings suggest that we can significantly improve the IAQ levels during the occupied periods by dynamically operating the HVAC systems based on both TVOC and CO_2 data in targeted zones.



(a) CO_2 concentration (b) TVOC concentration

Figure 2.4: Histograms of CO_2 and TVOC concentration in single-occupancy office C

2.4.1 Single-occupancy Offices

In this section, we evaluate the TVOC and CO_2 levels of three single-occupancy offices and assess how often these pollutants exceed the recommended thresholds. Figure 2.4 provides a histogram that depicts the frequency distribution of TVOC and CO_2 concentrations in Office C during working hours over a four-month data collection period. The data reveals that CO_2 levels ranged between 400 and 1000 ppm, while TVOC levels varied from 0 to 600 ppb in this specific office. Peaks in the histograms indicate that the most common concentration ranges for CO_2 and TVOC are 460-480 ppm and 100-125 ppb, respectively — both of which fall beneath their recommended levels. It also shows that CO_2 level never exceeded its recommended level in this room, while the red bars in the TVOC histogram indicated the occurrence of TVOC levels above the recommended level (6% of the time). we presented the histograms for TVOC and CO_2 concentrations of the other two single-occupancy offices in the appendix section (A.1and A.2). Tables 2.3 and 2.4 display the percentage of time these pollutants exceeded the recommended thresholds during working and off-hours, respectively. In these tables, the frequency of poor IAQ situations in the third row is sometimes lower than the frequency of TVOC levels exceeding their recommended limits in the second row. This difference occurs because instances where CO_2 levels have already triggered ventilation due to their high levels are excluded from the calculation of poor IAQ frequency when TVOC levels also exceed their recommended limits. Across the single-occupancy offices, the CO_2 concentration is below the identified moderate range during both working hours and off hours; however, this is not the case for the TVOC concentration. We can see that the TVOC levels are above the limit level at 6%, 3%, and 6% of the time in offices A, B, and C, respectively, during working hours.

Pollutant	Office A	Office B	Office C	Conference room	Hallway	Arena	Shared office space 1	Shared office space 2
CO_2 (%)	0	0	0	3	0	0	0	0
TVOC (%)	6	3	6	37	8	24	15	16
Poor IAQ (%)	6	3	6	34	8	24	15	16

Table 2.3: Percentage of time pollutants exceeded recommended levels and poor IAQ occurs during working hours

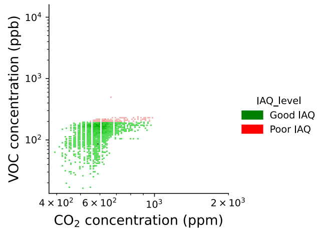
During off hours, single-occupancy offices A, B, and C experienced elevated TVOC

Pollutant	Office A	Office B	Office C	Conference room	Hallway	Arena	Shared office space 1	Shared office space 2
CO_2 (%)	0	0	0	2	0	0	0	0
TVOC (%)	33	18	19	30	15	36	30	22
Poor IAQ (%)	33	18	19	28	15	36	30	22

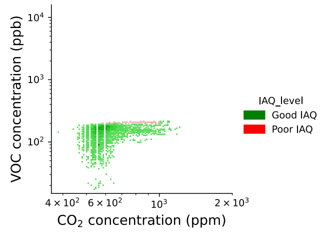
Table 2.4: Percentage of time pollutants exceeded recommended levels and poor IAQ occurs during Off hours

levels, with the accumulation surpassing recommended thresholds 33%, 18%, and 19% of the time, respectively. To obtain information about the poor IAQ situation in which air cleaning actions would not be taken, we need to assess the values of the two pollutants jointly. Fig. 2.5 (plot a-c) shows the joint plot of CO_2 and TVOC concentrations for single-occupancy offices, and Table 2.3 shows the percentage of time that the IAQ is in poor condition during working hours. Notably, instances where both CO_2 and TVOC levels surpassed the recommended limits are also marked in green in Fig. 2.5. The reason is that when CO_2 is above the recommended level, the ventilation system would be activated as they consider this pollutant as the only proxy of IAQ. Conversely, instances marked in red indicate periods where only TVOC levels are elevated, leading to their accumulation due to the absence of ventilation response, adversely affecting occupants' well-being and performance. The highest frequency of poor IAQ happens in offices A and C, which is 6% for both offices during working hours. This shows that the IAQ in these offices is poor 6% of the time during the normal weekday schedule in which the HVAC system is operating at a 2.6 ACH air exchange rate.

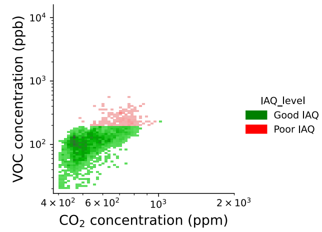
The results of our SVM model and motion sensor data (where they were available) show that during working hours, 11%, 7%, and 16% of the occupied time, there was



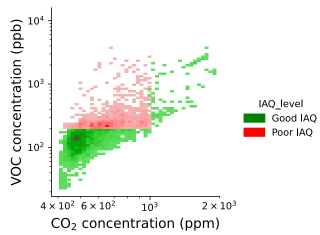
(a) Single-occupancy office A



(b) Single-occupancy office B



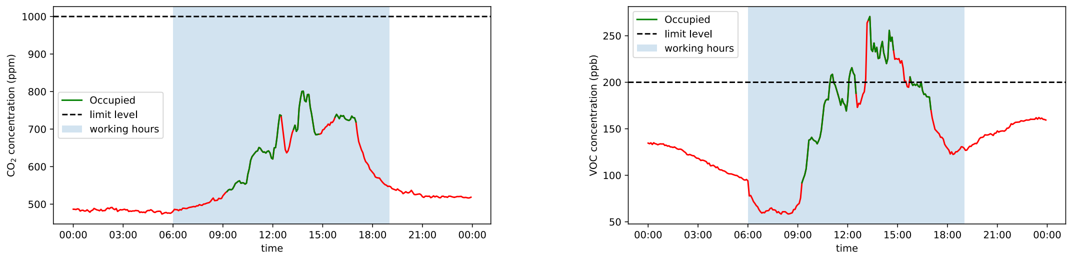
(c) Single-occupancy office C



(d) Conference room

Figure 2.5: IAQ situation in single-occupancy offices and the conference room during working hours

at least one occupant in the single-occupancy offices A, B, and C, respectively, while the IAQ was poor. These values correspond to a daily average exposure of 45, 20, and 40 minutes for each occupant in single-occupancy offices A, B, and C, respectively. Fig. 2.6 shows a time series plot of TVOC and CO_2 levels of a working day in a single-occupancy office with the occupied occasions. As shown in Figure 2.6, both TVOC and CO_2 levels exhibit a gradual increase following the arrival of an occupant. During the occupied time, the TVOC level can exceed the recommended moderate level, while the CO_2 level stays below the recommended level. Evaluating occupied periods on other days also showed the same pattern of TVOC and CO_2 concentration.

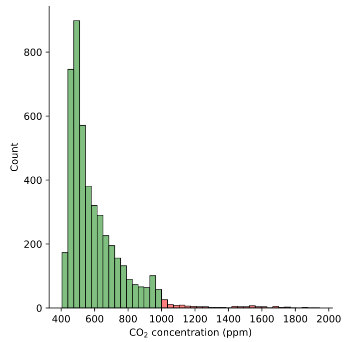


(a) CO_2 level　　(b) TVOC level

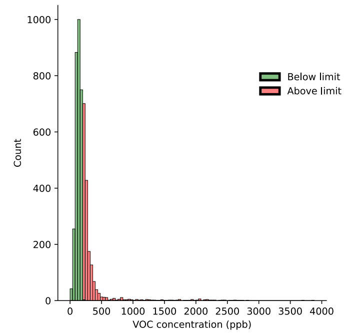
Figure 2.6: CO_2 and TVOC time series data during a working day in a single-occupancy office

2.4.2 Conference Room

Figure 2.7a and 2.7b present histograms depicting TVOC and CO_2 concentrations in the conference room, respectively. The histograms reveal that CO_2 levels ranged from 400 to 2000 ppm, while TVOC levels varied between 0 and 4000 ppb. Notably, the most frequent concentrations for CO_2 and TVOC were 480-520 ppm and 120-160 ppb, respectively, both of which fall below their respective recommended limits. The red bars in these plots signify instances where pollutant levels exceeded recommended guidelines. As detailed in Table 2.3, TVOC levels surpassed the recommended limit 37% of the time during working hours in the conference room, compared to just 3% for CO_2 levels. Figure 2.5 (plot d) illustrates that poor IAQ conditions prevailed 34% of the time during working hours—a concerning statistic given the room's frequent turnover. Utilizing both the room's calendar and motion sensors for occupancy detection, we discovered that during 71% of the occupied time in working hours—equivalent to 3 hours and 50 minutes—the room experienced poor IAQ conditions. Further, calendar data and other metrics suggest that the room typically accommodates more than one occupant when IAQ is in poor condition. These values showed significant periods in which occupants have meetings in the conference room, but the



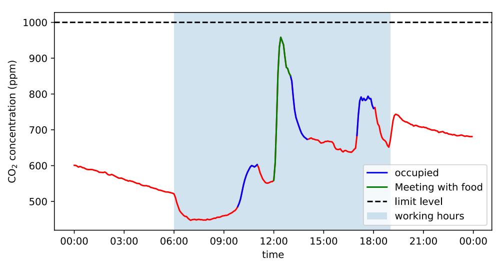
(a) CO_2 concentration



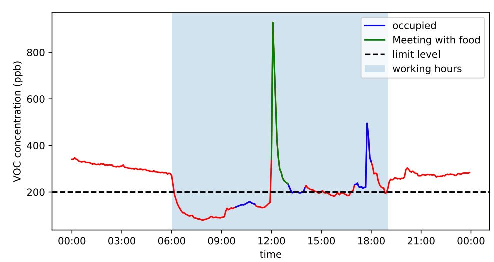
(b) TVOC concentration

Figure 2.7: Histograms of CO_2 and TVOC concentration in the conference room

HVAC system is not circulating enough air to improve the IAQ levels. Evaluating the poor IAQ during off hours shows that 28% of the time, the space has poor IAQ levels.



(a) CO_2 level



(b) TVOC level

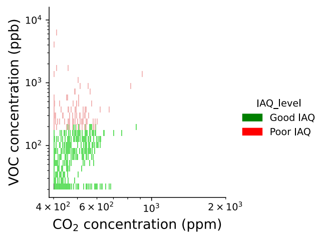
Figure 2.8: The average CO_2 and TVOC time series data during a working day for four weeks in the conference room

Figure 2.8 shows the average levels of TVOC and CO_2 for four representative Mondays during the data collection period, focusing specifically on days when recurring meetings were scheduled in the conference room. The number of people and their activities in the room varied across different meetings, leading to fluctuating levels of TVOC and CO_2 concentrations. For example, in the meeting at 12-1 PM, food was served every week, resulting in higher levels of TVOC. As illustrated in Figure 2.8, the presence of occupants led to increases in both TVOC and CO_2 levels. However, the CO_2 concentrations remained within acceptable range throughout the day, unlike TVOC levels, which frequently exceeded the recommended thresholds. Like the single-occupancy offices, the conference room also experienced elevated TVOC levels from 7 PM to 6 AM, which directly resulted from the reduced ventilation rates. During this period, the accumulation of TVOC occasionally resulted in poor IAQ. Similar patterns in fluctuations of CO_2 and TVOC levels are evident in the conference room across other days of the week.

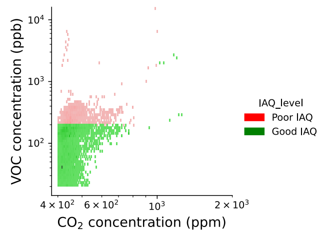
2.4.3 Open-space Areas

Fig. 2.2a shows the locations of four different open-space areas, including the shared office spaces occupied by graduate students. Fig. 2.9 shows the poor IAQ conditions in each of these four spaces. In this figure, the x-axis represents CO_2 concentration, while the y-axis reflects TVOC levels. Red dots in Fig. 2.9 mark instances in which TVOC concentrations exceeded the recommended threshold of 200 ppb, even as CO_2 levels remained below their advised limit of 1000 ppm.

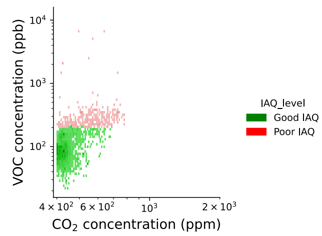
Table 2.3 quantifies the percentage of working hours during which poor IAQ was observed across these areas. Values in this table reveal that the poor IAQ situation happens more often near the kitchen and dining area. In this area, 24% of the time, the air quality was poor during working hours. It is important to note that occupants do not typically spend long periods near this area; therefore, their exposure to higher levels of TVOC in this space is limited. In shared office spaces 1 and 2, the IAQ levels were poor at 15% and 16% of the time during working hours, respectively. This area is where the occupants work at their desks for long periods. The entrance corridor, which benefits from regular influxes of fresh air, recorded the lowest incidence of poor



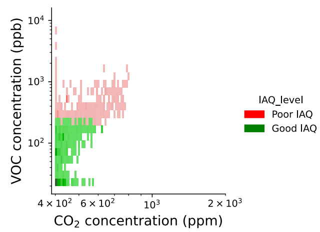
(a) Entrance Hallway



(b) Kitchen and arena



(c) shared office space 1

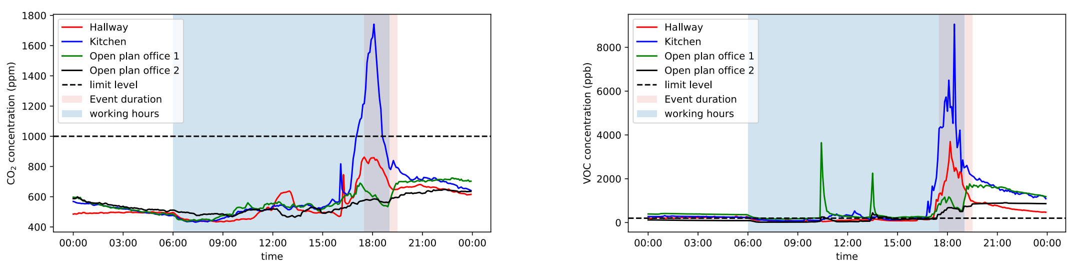


(d) shared office space 2

Figure 2.9: Poor IAQ situation of the open-space areas during working hours

IAQ at 8% during working hours. Our findings show that these values are at higher rates during the off hours when the HVAC system operates at a lower rate. During off hours, which include weekdays from 7 PM to 6 AM and all day on weekends, the TVOC levels in various areas increased due to reduced ventilation rates. This led to sporadic TVOC accumulation, exceeding recommended levels and resulting in poor IAQ. Specifically, the incidence of poor IAQ during these off hours was recorded at 15% in the hallway, 36% in the kitchen area, and 30% and 22% in shared office spaces 1 and 2, respectively. As shown in both Table 2.3 and 2.4, it's noteworthy that CO_2 levels never surpassed their recommended limits (happened less than 1% of the time only during the events in the arena), neither during working hours nor off-hours.

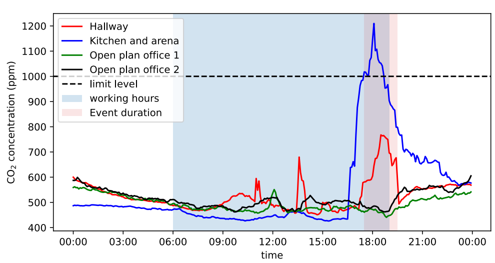
During the data collection period, the L-LL arena and kitchen area hosted three events and one large gathering, each beginning around 5:00 PM and ending around 7:30 PM. To assess pollutant levels during these occasions, we took the raw data and averaged it into five-minute intervals for both CO_2 and TVOC on the event days. Figure 2.10 displays the 24-hour average concentrations observed on those specific dates. Two of the events had an attendance of roughly 70-80 people, while the remaining two drew between 30-40 participants. Food was also served in all four events. For comparative purposes, Figure 2.11 isolates the pollutant levels from the smaller events (30-40 attendees). In these smaller gatherings, CO_2 peaked at 1200 ppm, and TVOC reached around 4000 ppb—levels that are appreciably lower than those observed in larger events, where CO_2 hit 1800 ppm, but TVOC soared to an alarming value around 8000 ppb, as shown in Figure 2.10. The average of the TVOC level in the event location was 4078 ppb during the event period.



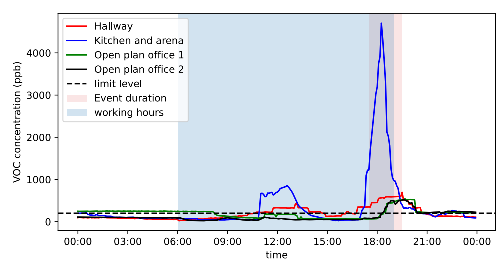
(a) CO_2 level　　(b) TVOC level

Figure 2.10: Average CO_2 and TVOC time series data during days of four social events in the arena

A significant observation shown in Fig. 2.10 is that the ventilation system failed to mitigate TVOC concentrations in any of the open-space areas by the end of the day. As previously mentioned, during off hours, there was an increase in TVOC levels in these areas, with accumulations sometimes exceeding recommended levels. Notably, following such events, TVOC concentrations remained above the recommended threshold until 6 AM, when the ventilation systems resumed operation with a higher ACH rate to cleanse the air of indoor pollutants. This is while CO_2 was above the standard level only in the arena for about 90 minutes during the events. As we can see in Fig. 2.10, TVOC concentrations initially spiked in the arena and gradually permeated other lab spaces, a pattern not seen with CO_2, which remained elevated for a much shorter duration and was confined mainly to the arena where the events were happening. Based on the preset schedule of the lab's ventilation system, the HVAC systems switched to a lower ventilation speed at 7:00 PM, regardless of the pollutant level or occupancy levels. This contributed to TVOC concentrations staying above the recommended levels for nearly six hours post-event across all lab spaces. As shown in Fig. 2.11, when the event is smaller with fewer attendees, the ventilation systems decreased the TVOC level to the recommended level about an hour after the event ended. As we can see, even in smaller events, the TVOC level increased all around the lab, despite the CO_2 level, which only surpassed the recommended range in the event area.



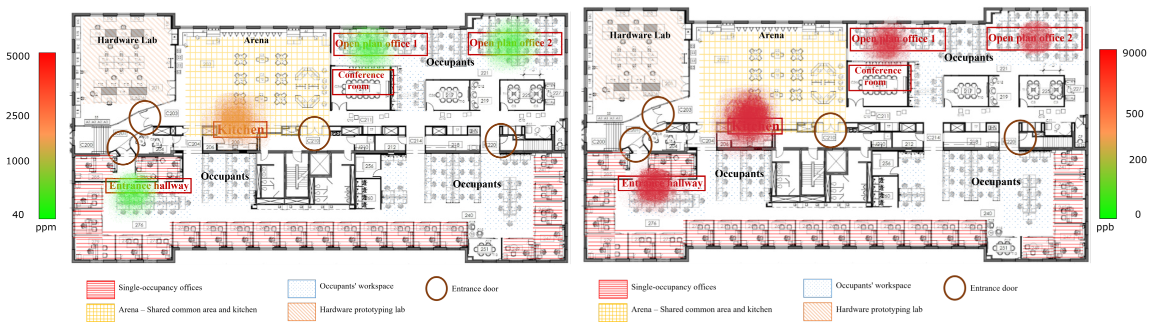
(a) CO_2 level



(b) TVOC level

Figure 2.11: Average CO_2 and TVOC time series data during days of two social events with approximate 30 attendees

Figure 2.12 provides an alternative visualization of pollutant levels and their distribution across various open-plan locations within the lab during these events. This heatmap illustrates the peak levels for both TVOC and CO_2 levels at each location. The least affected area by the spread of TVOC is shared office space 2, in which the TVOC level went up to 520 ppb in smaller events and 905 ppb in larger events. As indicated in Fig. 2.12, the CO_2 levels across different lab areas are comparatively less influenced by the events. This could likely be attributed to a lower concentration of



(a) CO_2 level distribution (b) TVOC level distribution

Figure 2.12: CO_2 and TVOC distribution over the lab during an event

people outside the main event area.

The entrance hallway is the most impacted area for CO_2 concentration, following the main event arena. In this hallway, CO_2 levels peaked at 862 ppm in larger events and 767 ppm in smaller events—levels that still fall below the recommended level as defined in this study. Notably, the entrance hallway and the arena are the only two areas where CO_2 levels exhibited an increase due to event activities. Even so, in the entrance hallway, the CO_2 concentrations remained below the recommended level. Upon examining the IAQ both during and after these events, it becomes evident that the current preset HVAC system operations are inadequate for addressing the dynamic changes in space utilization. Specifically, the system struggled to purify the air expediently even once the events, particularly the larger ones, ended.

2.5 Conclusions and Discussion

This study investigated the spatiotemporal variations of CO_2 and TVOC levels as primary indicators of poor IAQ over four months in a multi-functional institutional building. We examined the influence of occupant presence, activities, and HVAC system ventilation rates on pollutant concentrations in diverse spaces within the University of Virginia's living lab. CO_2 levels above 1000 ppm and TVOC levels exceeding 200 ppb were considered poor levels of these pollutants. Our analysis, conducted during both working and off-hours, assessed the frequency of poor IAQ events and the effectiveness of using CO_2 as the sole metric for IAQ and ventilation adjustments. We identified instances where poor IAQ conditions were not detected by the HVAC system due to its reliance on CO_2 levels alone, despite elevated TVOC concentrations.

The assessment of IAQ across various zones within our testbed reveals that poor IAQ conditions are relatively infrequent in single-occupancy offices during working hours. The highest value obtained in these offices for poor IAQ situation is 6% of working hours. Also, using the machine learning models for occupancy presence showed that in the same office, the IAQ was poor during 16% of the occupied time. These values were higher in the conference room with the occurrence of poor IAQ during 34% of the working hours and 71% during the occupied periods (during working hours). Evaluating this situation in open plan spaces showed that the highest occurrence of poor IAQ condition was in the kitchen and arena (24% of the time during working hours), where occupants eat their food and hold gatherings and events. Next in line were shared office spaces 1 and 2, showing poor IAQ for 15% and 16% of working hours, respectively. As these spaces are primarily occupied by graduate students who spend extended periods at their individual desks, enhancing ventilation rates in these spaces becomes particularly crucial. The poor IAQ situation only occurs 8% of the time during working hours in the entrance hallway. This area did not host many students at the time of data collection and is located near one of the main entrance doors where the occupants are in transit rather than stationary. During off hours, TVOC levels were elevated due to the lower ventilation rates, and as a result, every area, except the conference room, experienced a higher frequency of poor IAQ conditions compared to their working-hour counterparts.

Our analysis of pollutant levels during social events in the lab underscores the necessity of including TVOC as key factors in determining ventilation rates. Based on our results of evaluating different locations of the lab with different applications, occupancy levels, and patterns, we conclude that it is essential to consider TVOC in automating HVAC systems in areas with more occupants, such as conference rooms and shared office spaces. Our results show that a preset schedule of the HVAC operations and considering CO_2 as the only signal for ventilation rates is insufficient for cleaning indoor air of high occupancy areas. Also, dynamic operation logic with TVOC consideration for ventilation systems is needed to clean the indoor air during social events. This is while areas with fewer occupants, such as single-occupancy offices, are less prone to poor IAQ (up to 16% of occupancy period) even with the preset schedules of ventilation systems.

For controlling the HVAC operation, rather than considering only the values we obtained for poor IAQ frequency, we need to consider the exposure duration of the occupants to these compounds as well. For instance, although poor IAQ is more frequent in the kitchen area than in shared office spaces 1 and 2 —where graduate students typically work at their desks— the latter still demands close attention. This is because the potentially severe health impacts of long-term exposure to elevated TVOC levels are a concern for students who spend extended periods at their desks. In contrast, individuals who merely pass through the hallway or the kitchen experience shorter exposure times. Therefore, some spatial intelligence of how different spaces are utilized is also important for the future optimization of HVAC operations.

2.6 Summary of contributions

- Analyzed CO_2 and TVOC levels over a four-month period in real-world, naturalistic conditions to assess air quality dynamics.
- Demonstrated that reduced preset ventilation settings often result in TVOC accumulation, highlighting the need for more adaptive ventilation strategies.
- Identified the limitations of preset HVAC schedules in effectively managing air quality, particularly in high-occupancy environments.
- Established that occupants' actions and specific events have a greater influence on TVOC variations than on CO_2 changes, underscoring the need for more responsive ventilation systems.

Chapter 3

Detecting Occupancy status and level using indoor environmental factors

3.1 Abstract

Occupancy models can effectively optimize building-systems operations to prevent energy waste and achieve optimal indoor environmental quality. Previous research has relied on CO_2 sensors and vision-based techniques to determine occupancy patterns. Vision-based techniques provide highly accurate information; however, they are very intrusive. Therefore, motion or CO_2 sensors are widely adopted worldwide. Total volatile Organic Compounds (TVOC) are another pollutant originating from the occupants. However, a limited number of studies have evaluated the impact of occupants on the TVOC level. This study recorded continuous CO_2, TVOC, light, temperature, and humidity measurements in a 17,000 sqft open office space for around four months. Using different statistical models (e.g., SVM, K-Nearest Neighbors, and Random Forest), we evaluated which combination of environmental factors provides more accurate insights into the occupants' presence. Our preliminary results indicate that TVOC is a good indicator of occupancy detection in some cases. It is also con-

cluded that using a proper feature selection method can reduce the cost and energy of data collection without significantly impacting accuracy. Other than the occupancy status model, using 14 months of data from a conference room, we developed a CBLSTM model for detecting the number of occupants in the conference room, achieving a mean absolute error (MAE) of 2.04.

3.2 Introduction & background

Buildings are major consumers of energy in the world, accounting for approximately 40% of the total energy consumption [56]. Heating, ventilation, and air conditioning (HVAC) systems are considered as the highest contributors among building-related energy consumption sources and are responsible for approximately 48% of the total energy consumption in the buildings [57]. Consequently, HVAC systems can be one of the main targets to be considered when thinking of reducing energy consumption in buildings. Traditional HVAC systems operate based on the maximum design occupancy of a building during the occupied hours, which may cause an unnecessary increase in HVAC operations and ultimately increase energy consumption [58]. Meanwhile, reducing the ventilation rates has resulted in occupants' discomfort [59], as well as increasing the chance for spreading harmful particles and viruses, as we have learned more due to the COVID-19 pandemic.

In fact, traditional HVAC operations can lead to both energy waste and occupants' discomfort, highlighting the need for a change in their operations. To address this, having accurate information on building occupancy levels can help replace traditional HVAC operations with demand-response HVAC control [60, 61]. Demand-response HVAC control can significantly reduce the buildings' energy consumption during the occupied hours and off-hours by dynamically adjusting the air ventilation according to the occupancy status and patterns of specific zones while preventing overcooling or over-heating of vacant zones [62, 63]. Prior research has shown that by having access to accurate occupancy counts and patterns, building automation systems (BAS) can dynamically adjust and control the ventilation rates of the HVAC systems in different zones, resulting in up to 80% reduction in HVAC-related energy consumption [64]. Moreover, occupancy information is important for emergency evacuation [65], security management [66], and controlling lighting systems of buildings [67].

With the recent advancements in the Internet of Things (IoT) along with ubiquitous computing, we can accurately develop an occupancy model that can inform the operation of HVAC systems [68, 69]. Low-cost and easy-to-deploy non-intrusive indoor environmental quality (IEQ) sensors such as CO_2, temperature, humidity, and light sensors have become widely available in modern buildings and have shown to provide promising results in detecting occupancy counts and patterns [70, 71]. Since occupants directly influence the pollutants in indoor environments by their presence, activities, and the products they use (e.g., perfume), using IEQ sensors can greatly help detect occupants' presence [68].

Among the IEQ factors, CO_2 sensors are the most common modality that has been studied to identify occupancy estimation and detection. Although CO_2 has been shown to be effective for occupancy detection, the slow spread of CO_2 in the environment and the time that it takes for CO_2 to build up will cause the results to have a time delay from the real building occupancy. Another limitation of CO_2 sensors is that there are other factors, such as passive ventilation and sensor placement, that affect the CO_2 levels. As an example of the importance of CO_2 sensor placement, Pantelic et al. show that occupants have a personal CO_2 cloud, and if the CO_2 sensors are not adequately placed near the occupant's CO_2 cloud, their readings could be delayed, or incorrect [41]. As a result, the placement of these sensors is widely important and often not considered when such sensors are placed in indoor environments.

In addition to CO_2, other environmental factors such as temperature, humidity, and light have also been used for occupancy detection tasks. Candanedo and Feldheim used environmental features including CO_2, light, temperature, and humidity and compared three learning algorithms, namely linear discriminant analysis (LDA), classification and regression trees (CART), and random forest (RF), in their occupancy detection system. They showed that satisfactory results could be obtained by using proper feature selection and learning methods [68]. In another work, Kraipeerapun and Amornsamankul used stacking for multiclass classification and binary detection of occupancy through utilizing CO_2, light, temperature, and humidity levels as input features [72].

Over recent years, due to the availability of more reliable commercially available IoT devices, monitoring other IEQ factors has also gained more attention. For instance, recent studies have shown that occupants' presence can have a significant impact on the changes in total volatile organic compounds (TVOC). These changes could be a result of the presence of organic materials in indoor spaces, such as food, cleansers or disinfectants, and aerosol sprays. However, to the best of the authors' knowledge, no study has evaluated whether TVOC can provide similar or better occupancy-related information as other evaluated metrics such as CO_2, lighting, or temperature.

To overcome the mentioned limitations in previous research, this paper evaluates whether TVOC data is a reliable indicator of the occupancy status of single- and multi-occupied spaces. Furthermore, we evaluate whether using different environmental factors, including CO_2, TVOC, light, temperature, and humidity, can bring better insights into detecting occupancy presence. Different statistical models (i.e., SVM, K-Nearest Neighbors, and Random Forest) are used for this task. The Random Forest feature selection technique substitutes the manual selection of features that require expertise and may miss some important features. In addition to models for detecting the occupancy status, we have also developed deep neural networks to predict the number of occupants in a conference room using CO_2 and TVOC as the features.

3.3 Methodology and Results

3.3.1 Detecting occupancy status

The data collection for this study was conducted in Living Link Lab (L-LL) ("Living Link Lab" 2021) – a 17,000 sqft living lab located on the University of Virginia campus, which includes approximately 250 occupants. L-LL consists of 24 single-occupancy offices, four conference rooms, an arena space, and three large open-layout spaces (Figure 2.2a). This space is equipped with over 200 IEQ sensors, including temperature, humidity, light, motion, air quality (CO_2, TVOC, $PM_{2.5}$), and noise-level sensors. These sensors have been deployed over the past four years, and as a result, granular longitudinal data has been collected over this period.

For this section, we have chosen two single-occupancy offices (A and B) and one conference room, which included at least two motion sensors, light-level, temperature, humidity, and at least one air quality (CO_2 and TVOC) sensor. The selected time period for data analysis was from January 15, 2020, to April 30, 2020, which includes both pre-COVID-19, when space was at full occupancy, and post-COVID-19, when the offices were not occupied, and HVAC systems were operating as normal. This helps us have enough data points for both occupied and non-occupied situations in the space. Another step toward balancing our dataset is applying the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class by identifying k-nearest neighbors based on Euclidean distance in the feature space. By employing SMOTE and carefully selecting the time periods, we ensured the dataset used for analysis was balanced and representative. Since the space is considered a living lab, occupancy schedules were not pre-defined for single-occupied offices, and the occupants in those rooms were living their normal lives. Ground truth for occupancy status was obtained using two motion sensors as well as occupants' calendar information, indicating when they were in and out of their office. Similarly, motion sensors and the conference room's shared calendar were utilized to obtain the ground truth for occupied versus non-occupied times in the conference room.

In this phase, occupancy modeling is treated as a binary classification with the possible outcome of 0 when there is no occupant and 1 when there are one or multiple occupants in the space. We adopted five machine learning models, including support vector machine (SVM), Gaussian Naive Bayes (GNB), logistic regression (LGR), random forest classifier (RFC), and k-nearest neighbor (KNN) to identify which would fit the data better. Light, temperature, humidity, TVOC, and CO_2 levels, which were collected during the identified time frame, were the input features for these algorithms. In the two single occupancy offices, two different CO_2 levels are considered, one with installing a sensor close to the occupants (CO_2_inhale) and the other placed in the background area (CO_2_bg) to evaluate the impact of CO_2 sensor placement.

For the SVM technique, C-Support Vector Classification with a kernel function is used. We used linear and non-linear kernels to train the model and chose the one that outperformed. Based on the results, when we use more than two features for the occupancy detection task, using a Gaussian kernel (as a non-linear kernel) leads to higher accuracy. This paper uses the radial basis function (RBF) as the Gaussian kernel. When training an SVM with the RBF kernel, two parameters should be considered, the parameters C and gamma. Parameter C is common in all SVM kernels and trades of misclassification of training examples against the simplicity of the decision surface. When C has a low value, it smooths the decision surface, while a high C aims to classify all training examples correctly. For tuning the parameter C, we have used the cross-validation method, and for the parameter gamma in the RBF, which affects the decision boundary's flexibility, the equation 3.1 is used, in which n_{features} is the number of features and X is the input (training data).

$$\gamma = \frac{1}{n_{\text{features}} \cdot \text{Var}(X)} \tag{3.1}$$

Additionally, we tested the LGR model that can be used when the output of a problem is categorical, which in this case, the occupancy detection is in a binary form. This paper uses the L-BFGS optimizer with L2 regularization for the LGR algorithm. The C parameter, which is the inverse of regularization strength, is set to 10. This parameter is similar to the one in the SVM algorithm, and the smaller values specify stronger regularization.

The neighbors-based classification was tested, a type of instance-based learning that does not attempt to construct a general internal model but instead attempts to store instances of the training data. In this paper, we used the KNN model in which the classification is implemented by using the majority vote of the nearest neighbors for each data point. We used uniform weights for this model. Parameter k in this model is the number of neighbors to use for queries and in general, larger values of k suppress the effects of noise while making the classification boundaries less district. We have evaluated the model performance with k between 0 and 10 and set the parameter to 5, which outperforms other values. The distance metric parameter is set to Minkowski with a power parameter of 2, which is equal to the Euclidean distance.

Furthermore, the Naive Bayes (NB) method was tested which is a set of supervised learning algorithms based on Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Bayes' theorem states the following relationship:

$$P\left(y \mid x_{1}, \ldots, x_{n}\right)=\frac{P(y) P\left(x_{1}, \ldots, x_{n} \mid y\right)}{P\left(x_{1}, \ldots, x_{n}\right)} \tag{3.2}$$

in which y is the category variable and x_1 through x_n are the feature vectors. Based on the strong assumption in NB about the independence of all features, we used the Maximum A Posteriori (MAP) estimation the Gaussian NB (GNB) was considered for this classification task.

Results showed that both TVOC and CO_2 levels significantly change upon the presence of occupants. For evaluating each of the mentioned models, we have used individual environmental features and a combination of them to evaluate the impact of the features on the accuracy of each model. Table 1 shows the results of using the predictive models for occupancy detection in all three offices. As shown in the results across all predictive models, for single occupancy offices, using CO_2_bg (background CO_2) decreases the accuracy by around 10% compared to the CO_2_inhale. As we can see in the results of both single-occupancy offices, CO_2 is a good indicator of oc-

Office A/ Office B/ Office C					
Feature	SVM	GNB	LGR	RFC	KNN
CO_2_inhale	84/88/-	78/87/-	79/87/-	79/87/-	76/81/-
CO_2_bg	75/75/64	64/72/65	64/73/64	67/70/69	66/66/61
TVOC	76/63/56	71/50/30	66/46/39	68/44/50	68/52/56
Light	76/-/82	56/-/80	61/-/81	70/-/79	66/-/78
Temperature	76/63/61	55/58/69	57/58/69	57/56/70	61/52/62
Humidity	76/63/64	45/45/61	42/47/60	50/48/51	61/50/60
CO_2 + TVOC	85/88/74	79/86/69	83/87/68	84/86/71	86/85/65
CO_2 + TVOC + light	85/-/82	80/-/81	83/-/81	84/-/80	86/-/70

Table 3.1: ML models accuracy in predicting the occupancy binary status in different locations

cupant's presence in both offices, but TVOC is a good indicator only in office A. This can be because people emit CO_2 by breathing while their activities and products they use/wear are sources of TVOC. Therefore, the performance of TVOC as an indicator of the presence of the occupants in the space is greatly influenced by the occupants' characteristics and profile.

In office A, TVOC is a better indicator of occupancy than the background CO_2. The conference room does not get any sunlight and the best feature showing the occupancy status is light since the lighting is controlled by the occupants. In the Link Lab, the control of occupants on the temperature is limited and they have no control over humidity, and as we can see, the results show that these features are not a good indicator for detecting occupancy presence in all three rooms and they have been removed by RF feature selection method. In the conference room, we can see that CO_2 and TVOC individually are not good indicators of occupancy; however, when combined, they improved the models' performance by up to 10%.

These modelings showed that TVOC could sometimes work better than background

CO_2 for detecting the occupancy level and can be used combined with CO_2 to improve the occupancy predictions. The results also showed that the location of CO_2 sensors is significantly important in detecting the occupants' presence. The performance of TVOC in detecting occupancy can depend on the occupants' activities and the products they use, as these are the ways by which occupants emit TVOC.

3.3.2 Detecting occupancy level

While binary occupancy status gives good information on how the building is being used, occupancy level (or count) information can be more helpful in spaces with high capacity (such as conference rooms). In these spaces, occupancy level information provides more granular insights into how a building is used compared to simple binary occupancy status. Also, when we only have the binary occupancy status, it is difficult to estimate occupants' performance loss, as we do not know how many people are affected by the poor IAQ. Information on the occupancy level can also help in managing HVAC operations. Ventilation rates can be reduced in areas with lower occupancy levels without negatively impacting air quality. Therefore, by knowing the occupancy level of different building areas, HVAC systems can adjust ventilation rates to better match the occupants' needs. This can save energy and reduce the load on the HVAC system.

For the previous section on detecting the occupants' presence, we used motion sensors to gain the ground truth of the target variable. There are some drawbacks to using motion sensors for occupancy detection. While these sensors provide good insight into the binary status of occupancy of the space, they can have false positives due to being triggered by non-human movements. Especially in the L-LL, where many doors are made of glass, these sensors can be triggered by the movements of objects outside the target space. The other concern regarding the motion sensors is the blind spot. These sensors can miss movements in certain areas based on their location and should be placed where they can catch most of the occupants' movements. Therefore, sometimes, these sensors can detect occupancy when there is none and can miss the occupancy even though someone may be present.

In the second phase of this study, to tackle these limitations and both improve the ground truth accuracy of binary occupancy status and gain information on the occupancy level, in the conference room, we installed an occupancy counter which is installed in the room. The sensor we are using uses a technology called infrared time-of-flight to anonymously detect and measure the movement of the occupants. This technology ensures privacy is protected and no personally identifiable information is collected. This sensor can count the number of occupants in the space with an accuracy of 98%. It has multi-directional counting and height filter capabilities, making it a decent sensor for our target location.

For this study, we installed 19 Awair sensors all over the conference room to fully detect the IEQ conditions of the room. Figure 3.1 shows the IAQ and occupancy counter sensor and their location in the conference room. We had this installment from June 2022 until August 2023 (14 months). All data (IAQ and occupancy data) were resampled into one-minute intervals by getting the mean value. Upon analyzing the collected data, we observed that all sensor readings were remarkably consistent, deviating by only 3-5% from each other. Utilizing both correlation analysis and random forest feature selection techniques, we identified one sensor (shown in the red circle in Fig. 3.1) as the most representative of the collective locations. This particular sensor's significance was further underscored by its strategic positioning, away from the direct influence of any ventilation vents, ensuring that its readings were not artificially impacted. Consequently, this location was selected for the experiment, balancing both statistical rigor and practical considerations in environmental monitoring. In the next step, for predicting the occupancy status, we trained the same models as the previous study (SVM (with both linear and RBF kernel), GNB, LGR, RFC, and KNN models) to evaluate how improving the ground truth data results in an improvement in the prediction. Therefore, the target variable is the binary occupancy status, and the features were TVOC and CO_2 data collected by the Awair sensors.

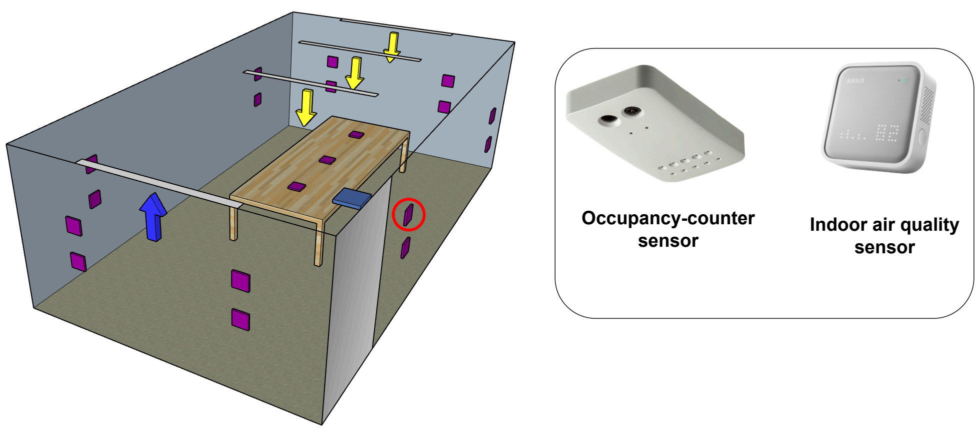


Figure 3.1: Left: sensors arrangement in the conference room. Purple cubes represent IAQ sensors, the blue cube represents the occupancy counter sensor, the red box is the selected location of the IAQ sensor for the data collection, and the yellow and blue arrows represent air inlet and outlet vents, respectively. Right: IAQ sensor and the occupancy counter sensor

Table 3.2 shows the results of ML models trained to predict the binary occupancy status using CO_2 and TVOC data. As we can see, the KNN, XGBoost, and RFC models' results significantly improved by improving the ground truth of the target variable, and their accuracies are all above 95%.

Other than improving the accuracy of binary occupancy status ground truth, the oc-

Evaluation Metric	SVM (linear)	SVM (RBF)	GNB	LGR	RFC	KNN	XGBoost
Accuracy	64	70	60	62	94	99	98
Precision	68	77	73	65	96	99	98
Recall	52	57	30	51	93	98	98

Table 3.2: ML models performance for predicting the binary occupancy status in the conference room using the occupancy counter sensor as the ground truth for the target variable

cupancy counter sensor is also used to gain information on the occupancy level. As the next step, we used ML models to predict the number of occupants by using the IAQ metrics as the feature variables and the occupancy level gathered from the occupancy counter sensor as the ground truth. In this phase, we employed both ML models and deep learning networks to ascertain their comparative effectiveness. Our analysis revealed that the Convolutional Bidirectional Long Short-Term Memory (CBLSTM) network significantly outperformed the traditional machine learning models previously utilized for binary occupancy status prediction. This approach begins with the raw CO_2 and TVOC sensor data being input into a convolutional network, which is structured around a convolutional layer paired with a max pooling layer. The convolutional layer efficiently processes the sequential data through a sliding filter window, extracting local features, while the max pooling layer is tasked with identifying and isolating the most discriminative of these features. This pooling layer highlights the essential characteristics and streamlines the model by reducing the overall number of features, consequently decreasing the number of parameters required [73]. This methodology enhances the model's efficiency and accuracy in predicting the precise number of occupants within a room.

Subsequently, the features extracted from the convolution network were input into a BLSTM network for further analysis. The LSTM network is distinguished by its proficiency in learning long-term dependencies within the data, a notable advancement over traditional Recurrent Neural Network (RNN) models [74]. Traditional RNNs, which rely on gradient backpropagation for training, often struggle with gradient vanishing and exploding problems. In contrast, the LSTM architecture incorporates specialized gates within its memory cells to regulate the flow of information, enabling it to preserve long-term dependencies effectively [74]. Our exploration of both unidirectional and bidirectional LSTM configurations revealed a critical insight that solely accounting for past temporal dependencies was insufficient for optimal model performance. The integration of future contextual information significantly enhanced the model's performance. Consequently, we opted for the BLSTM network as our final model due to its dual-direction processing capability, which systematically incorporates both past and future data contexts, offering a more comprehensive understanding of temporal dependencies.

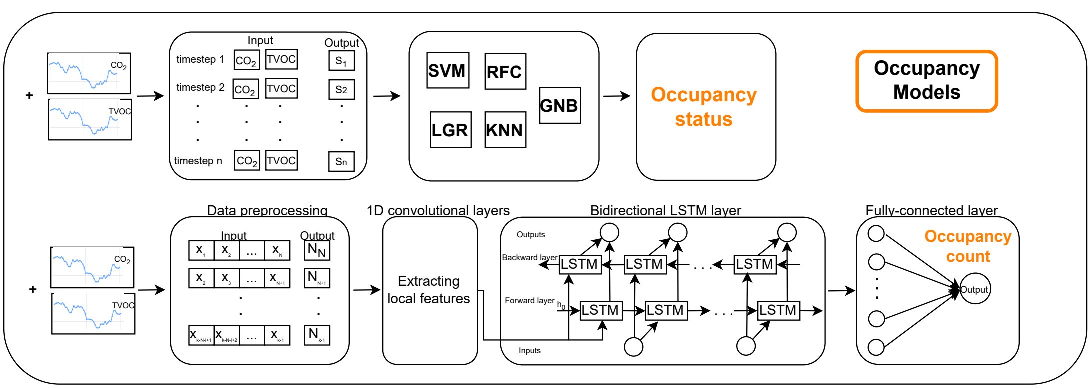


Figure 3.2: Overview of the machine learning models developed for occupancy status and occupancy counter prediction

Following that, the dense (fully connected) layers get the outputs from the BLSTM layer. The real occupancy count prediction is made in these layers, which also carry out the transition of the BLSTM output to the intended output shape. In order to translate features learned by the BLSTMs to the expected output, the fully connected layer mixes those features in nonlinear ways (considering the activation function). The occupancy count in the conference room is derived from the patterns the model has discovered in the CO_2 and TVOC data. We utilized Mean Absolute Error (MAE) as the loss function, and the outcomes revealed an MAE of 2.04.

3.4 Summary of contributions

- Predicting occupancy levels as binary status in single occupancy and the conference rooms
- Identifying the important features for detecting occupancy levels
- Predicting occupant(s) exposure to poor air quality in single occupancy offices and conference rooms
- Improving the models' performance on the binary occupancy detection task by improving the target ground truth accuracy by using an occupancy counter sensor instead of the motion sensor
- Predicting the level or the number of occupants in the conference room by using IAQ metrics as the feature

Chapter 4

Time series forecasting of the IAQ metrics

4.1 Abstract

Forecasting indoor air quality (IAQ) metrics is crucial for optimizing ventilation systems and ensuring a healthy indoor environment. This chapter investigates the application of both statistical and deep learning models to forecast IAQ metrics, using nine months of carbon dioxide (CO_2) and total volatile organic compounds (TVOC) data collected from a conference room. Linear models, such as Autoregressive Moving Average (ARMA), are compared with more advanced neural networks like Long Short-Term Memory (LSTM) and Convolutional Neural Network-LSTM (CNN-LSTM) combinations. Although ARMA models have been shown to be effective in predicting air pollution levels, their inability to account for nonlinear behaviors necessitated the adoption of deep networks for enhanced forecasting accuracy. The proposed CNN-LSTM framework leverages 1D convolution layers to extract short-term features from the data, while bidirectional LSTM layers capture long-term temporal dependencies. The models are trained to predict pollutant levels up to 30 minutes in advance, with results showing a significant improvement in prediction accuracy for the deep models compared to linear approaches. The best-performing CNN-LSTM

model achieved a mean absolute error (MAE) of 9.87 ppm for CO_2 and 7.13 ppb for TVOC forecasting. This chapter illustrates the capability of deep learning models to enhance the precision of indoor air quality forecasts, facilitating the adaptive management of HVAC systems to optimize occupant comfort and reduce energy consumption in building settings.

4.2 Introduction & background

Many researchers [75, 76, 77] have focused on predicting and forecasting thermal comfort-related parameters such as indoor temperature and humidity. For instance, Mustafaraj et al. [77] discussed using a linear parametric autoregressive model with external inputs (ARX) and a neural network-based nonlinear autoregressive model with external inputs (NNARX) to forecast room temperature and relative humidity in an open office. Features used for training the model were external and internal climate data recorded over three months, and target variables were forecasted from 30 minutes to 3 hours ahead. The study found that both models effectively predicted temperature and humidity, with the nonlinear neural network model providing better accuracy than the linear model. The study suggests that these models can be used to optimize HVAC systems to improve occupants' thermal comfort and energy efficiency in commercial buildings.

Also, suitable and effective forecasting tools were used for forecasting air quality metrics [78]. Generally, evaluating the air quality data by using statistical, ML, and deep neural models does not require an in-depth understanding of the dynamic and chemical processes between air contamination levels and building-related information. Statistical models make connections between the variables using probability and statistical averages and can gain acceptable accuracies in forecasting concentration levels of air pollutants. However, the behaviors of air pollutants could be complex and highly nonlinear. This is where machine learning models and artificial neural networks (ANN) can be helpful. Regarding outdoor air quality, due to the increase in air pollutants and deterioration of air quality, many predictive models were used to forecast the air quality metrics. However, the IAQ remains understudied in terms of the application of forecasting models.

As an example of using predictive models for outdoor air quality metrics, Kumar et al. [79] used the stationary stochastic ARMA/ARIMA (Autoregressive Moving (Integrated) Average) model to forecast the concentration of daily mean ambient air pollutants (O_3, CO, NO, and NO_2) at an urban traffic site in India. ARMA(0,1), ARIMA(0,1,3), ARIMA(1,1,.2), and ARIMA(3,1,3) were chosen for O3, CO, NO, and NO_2 as the best predictors respectively. For 20 out of sample forecasts, one step (i.e., one day) ahead MAPE (mean absolute percentage error) for CO, NO_2, NO, and O_3 was 13.6, 12.1, 21.8, and 24.1%, respectively.

As mentioned, the IAQ is also very important and impacts occupants' well-being besides outdoor air quality and indoor temperature and humidity. The impacts of poor IAQ on the occupants differ according to many factors, such as the pollutants' level and exposure time. Several studies indicate that exposure to increased levels of these pollutants for a short period may result in breathing difficulties and eye irritation and could adversely affect pulmonary and cardiovascular health. However, extended exposure to these pollutants may cause harm to the respiratory, reproductive, neurological, and immune systems, as well as an increased risk of cancer [80, 81]. Regarding improving the IAQ, using time series models to forecast the IAQ metrics ahead of time can be very helpful in taking precautionary measures. Using these models would help minimize the occupants' exposure to poor IAQ by early detection of the situation, resulting in improved comfort and productivity in indoor environments. These models can help optimize the HVAC systems' performance by forecasting changes in indoor air pollutant levels and adjusting the system accordingly. For example, if a time series model predicts that TVOC levels will rise, the HVAC system can increase the ventilation rate to prevent the buildup of harmful air pollutants. This control logic for HVAC management not only results in improving occupants' comfort and productivity but also saves energy. Therefore, using time series models to forecast IAQ metrics is a valuable tool for HVAC management, allowing building managers to optimize system performance, reduce energy consumption, and maintain a comfortable and healthy indoor environment for occupants.

As an example of forecasting IAQ metrics, Krati et al. [82] have created an IoT sensing system to monitor and analyze the time variation of carbon dioxide in a university classroom. In their study, they only considered CO_2 level as the indicator of IAQ and developed models to forecast the CO_2 build-up in the classroom. The results showed that a significant increase in CO_2 concentration happened when the classroom was fully occupied. Comparing the models' performance showed that the hybrid model (combination of linear and non-linear models) performs the best in capturing both linear and nonlinear characteristics of the CO_2 data.

As mentioned, linear models such as ARIMA have been very useful in many applications, including forecasting air quality metrics. However, for several reasons, relying only on linear models might be insufficient to forecast indoor air quality metrics, such as CO_2 and TVOC levels. Firstly, indoor air quality pollutants exhibit complex and nonlinear dynamics, which linear models may not accurately capture. On the other hand, deep learning models are designed to learn complex nonlinear relationships between input and output data, making them potentially more suitable for IAQ metrics forecasting.

Secondly, indoor air pollutant levels are influenced by various factors (such as events in the area, number of occupants, and outdoor weather conditions) with complex interactions that evolve over time. These complex relationships and interactions can be challenging to model with linear models, which are typically based on assumptions of linearity and stationarity. Deep learning models can adapt to changing patterns and capture the dynamic relationships between the input and output data, making them more flexible and able to handle the complex interactions involved in IAQ metrics' forecasting.

Thirdly, deep networks can also incorporate other data sources to improve the accuracy of the IAQ metrics' forecasting. This is while incorporating additional data sources is not easily applicable in linear models such as ARIMA, as they typically rely solely on the past values of the time series. The application of this point in IAQ forecasting is that we can use different sources as input rather than only relying on the history of the target variable as the feature. Overall, while linear models such as ARIMA are useful for time series forecasting, using deep learning models can improve the accuracy of forecasting indoor air quality metrics due to the complex and nonlinear dynamics involved.

LSTMs are a type of recurrent neural network (RNN) that have been widely used for preprocessing and forecasting time series data in many applications. These models can learn the temporal dependencies in the time series data and accurately predict future values of the time series. To mention some, LSTMs have been used in financial prediction tasks such as stock price prediction, traffic predictions, weather prediction, and anomaly detection in time series data. LSTM has also been used in natural language processing (NLP) tasks, such as speech recognition and language translation. In these applications, LSTMs are used to model the temporal dependencies between words and sentences in the input text [83, 84].

Convolutional neural networks (CNNs) are a type of neural network used in deep learning for processing and analyzing visual data, such as images and videos. Convolutional layers can learn to extract features from the input data by convolving a set of learned filters with the input data. Some applications of the CNNs are object recognition and classification with the goal of identifying the presence of specific objects in an image, object detection with the goal of identifying the location of objects within an image, image segmentation to partition the image into different regions, and visual tracking to track the movement of objects within a sequence of images or videos. 1D convolutional layers are a type of convolutional layer that is specifically designed for processing one-dimensional sequential data, such as time-series data, audio signals, and text data. Some applications of the 1D convolutional layers are speech recognition, stock price prediction, text classifications, and signal processing. Overall, 1D convolutional layers are a powerful tool to process and analyze a wide range of time series data [85, 86].

Other than the separate applications of these two networks, there are different applications that use frameworks of a combination of CNN and LSTM networks. These applications focused more on multimodal projects involving both computer vision and NLP. For example, this framework has been widely used in visual question answering (VQA), Image captioning, and video analysis. For example, Yu et al. [87] proposed a multi-level attention network that uses both CNN and LSTM models for visual question answering. The visual features from the input image were extracted using CNN, and then region-based middle-level outputs from CNN were encoded into spatially embedded representations by a bidirectional recurrent neural network (LSTM). I also have used a framework of CNN and LSTM for generating descriptive sentences for video clips in one of my course projects [88]. In this project, we used CNNs (ResNet152) for the video encoding task, and for text decoding, we used the LSTM model. In our framework, the video encoder extracted feature vectors from the video frames and passed the output to the LSTM models to generate the captions for the corresponding video.

The combination of CNN and LSTM models can also be useful in forecasting tasks of time series data, such as IAQ metrics. As mentioned, CNN layers are commonly used for feature extraction in analyzing images and videos, but 1D convolutional layers can also be used for feature extraction in time series data. LSTM networks, on the other hand, are designed to capture long-term temporal dependencies in sequential data. So, the combination of CNN and LSTM networks can capture both short-term patterns (such as human occupancy, ventilation, and temperature effects) and long-term dependencies (such as seasonal patterns and time of day) in the data. This framework can help improve the accuracy of IAQ metrics' forecasting and provide better insights into the complex relationships of these metrics with the space's other conditions.

In this work, we used nine months of CO_2 and TVOC data collected from a conference room to train time series models to forecast these pollutants' concentration in the future. We will train linear statistical time series (such as ARMA) and deep Networks (LSTM and a combination of CNN and LSTM models) and compare their performance.

4.3 Methodology and results

4.3.1 Statistical models

In working with time series data, the first step is to observe the data carefully and implement some tests to make sure the data is stationary. In time series analysis, data stationarity refers to the data's statistical properties (such as mean, variance, and auto-correlation structure) being constant over time. In other words, the statistical properties of the stationary time series data remain the same irrespective of when the data is observed. This is an important step in time series data analysis because many statistical models and techniques assume that the underlying data is stationary. So, feeding non-stationary data to the models can lead to inaccurate or misleading results in modeling, forecasting, and hypothesis testing. There are several ways to check for stationarity in time series data, such as Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), and Variance ratio test.

ADF test is a unit root test that checks whether a time series has a unit root, indicating that the series is non-stationary. The null hypothesis of the ADF test is that the time series data has a unit root, while the alternative hypothesis is that the series is stationary. The test statistic for the ADF test is based on the regression of the first difference of the series on its lagged values, and the significance level is usually set to 5% or 1%. The KPSS test is another commonly used method to test for stationarity, which is a non-parametric test. This test checks for a trend in the time series data. The null hypothesis of the KPSS test is that the series is stationary, while the alternative hypothesis is that the series has a unit root, indicating non-stationarity. The test statistic for the KPSS test is based on the sum of squared deviations of the series from its trend line, and the significance level is also usually set to 5% or 1%. While both of these tests check for the stationarity of the time series data, they evaluate different aspects of stationarity. Therefore, getting different results from these two tests on the same dataset is possible.

We implemented both tests on the TVOC and CO_2 dataset of the conference room. The results of the ADF test indicated that p-value < 0.001, which rejects the null hypothesis of this test and indicates the data set as stationary. However, the result of the KPSS test indicates that the data set is not stationary by rejecting the null hypothesis with a p-value of 0.001. ADF and KPSS tests are complementary and should be used together to gain a complete understanding of the stationarity of the data. Based on the results from these two tests, visual signs, and the nature of our data, it seems that the data is not stationary, and we solved it by taking a first-order difference in the data. After pre-processing the data, we need to find the best model that fits our data. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are both used as model selection criteria in statistics. The main difference between these two methods is in how they penalize the model's complexity. The AIC is defined as $\mathrm{AIC} = 2k - 2\ln(\hat{L})$, where k is the number of parameters in the model, ln is the natural logarithm, and $\hat{L}$ represents the maximum likelihood. The AIC favors models that have a good fit to the data but penalizes models that are too complex, as measured by the number of parameters.

On the other hand, the BIC is defined as $\mathrm{BIC} = k\ln(n) - 2\ln(\widehat{L})$. , where n is the number of observations, ln is the natural logarithm, and $\hat{L}$ represents the maximum likelihood. The BIC also favors models that fit the data well, but it penalizes complexity more heavily than AIC, as measured by the number of parameters. Therefore, BIC selects the simplest model that provides an adequate fit to the data, while AIC tries to maximize the goodness-of-fit of the model while taking into account the complexity of the model. I used AIC to select the order of the ARIMA model that fits the data better for the forecasting task. Using AIC indicates that the best model that fits the data is ARMA(3, 2), with an AIC value of 84370.783.

4.3.2 Deep neural network models

To compare different models' performances, we also considered using deep neural networks to see if these models would improve the results. LSTM networks are a powerful tool in forecasting time series data due to their unique structure, which can link previous information to current tasks. The LSTM network is distinguished by its proficiency in learning long-term dependencies within the data, a notable advancement over traditional Recurrent Neural Network (RNN) models [74]. Traditional RNNs, which rely on gradient backpropagation for training, often struggle with gradient vanishing and exploding problems. In contrast, the LSTM architecture incorporates specialized gates within its memory cells to regulate the flow of information, enabling it to preserve long-term dependencies effectively [74].

A unidirectional LSTM network processes a sequence of input data in only one direction, typically from the past to the future. This traditional network only has access to past information at any given time step and cannot use future information to make its predictions. However, in a bidirectional LSTM network, the input sequence is processed both forward and backward through time by two separate RNNs, with their hidden states concatenated at each time step. This allows the network to capture dependencies in both directions and, therefore, can potentially improve the performance of the model in forecasting tasks. The bidirectional LSTMs would have more information, which can help them to better capture the long-term dependencies in the time series data. Additionally, bidirectional LSTMs can also help reduce the effect of vanishing and exploding gradients, a common problem in training deep neural networks. By processing the input sequence in both directions, the gradients can flow more easily through the network, leading to better training and more accurate forecasts. Due to these reasons, to better capture the long-term dependencies in the IAQ metrics, we used bidirectional LSTM layers in our proposed network.

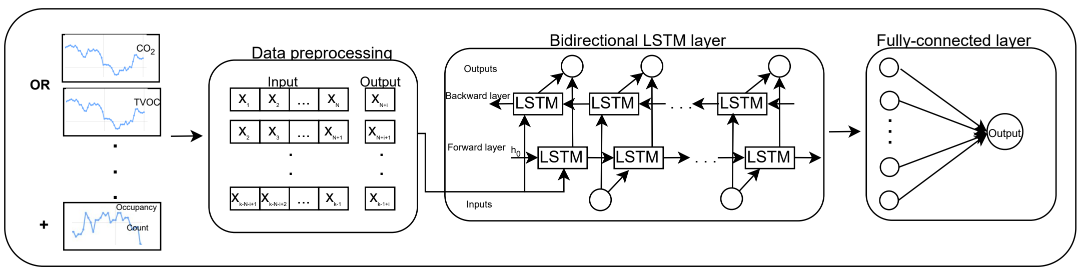


Figure 4.1: Proposed deep structure for IAQ metrics forecasting; N: sequence length, K: data size, i: time steps of forecasting in the future

As shown in Figure 4.1, the first step of the forecasting task is data preparation. The main steps of preparing the data for this time series forecasting tasks are normalization, splitting, and sequencing of the data. In general, deep learning models require data to be normalized or scaled to ensure that the optimization process during training is more efficient and effective. Normalizing the data can help gradient descent converge more quickly and efficiently since it helps keep the scale of the inputs consistent. The other benefits of normalizing the data are more effective activation functions, preventing vanishing/gradient descent, and acting as a form of regularization. In the second step, the data was split into training and testing sets without shuffling the data so as not to lose the dependencies of the sequential data. The last step of data preparation for time series forecasting is creating sequential data with the desired length (N) to be fed into the first layer.

In the next layer, the output of the pre-processing step was fed as the input of the LSTM layer, and the first hidden state was defined as a zero vector with a decent size. In this layer, the BLSTM captures the long-term dependencies of the data. The dropout layer is set in the LSTM layer, which is applied to the output of all LSTM layers except the final layer. Setting the dropout argument in the LSTM layer helps prevent overfitting the model by randomly dropping out (setting to zero) some of the units in the network during training. First, we applied dropout within the LSTM layer to regularize the recurrent connections across timesteps, reducing reliance on specific neurons and improving robustness in temporal feature learning. Second, we added a separate dropout layer after the LSTM outputs to regularize the feature representations before passing them to the fully connected layer. By randomly deactivating a fraction of neurons during training, this dual application of dropout helps mitigate overfitting, ensuring the model can generalize effectively to unseen data while maintaining its predictive accuracy.

The final output of the LSTM layer is passed to the fully connected layer without dropout so that the model can use all of the available information to make a prediction. In the last layer (which is a fully connected one), by using equation 4.1, the model outputs the forecast of the desired metric (for example, CO_2 level) in the desired time step (in the future). In this formula, the weight vector w contains the learnable parameters of the fully connected layer and is initialized randomly during training. The bias vector b is also learnable but is typically initialized to zeros. Vector x is the input of the fully connected layer (output of the LSTM layer), and y is the prediction (output). In this final step, the model forecasts the levels of pollutants (CO_2 and TVOC) up to 30 minutes in advance.

In order to further enhance the generalization of the LSTM-based forecasting model and reduce the issues of overfitting, we have combined dropout with the L2 regularization technique. L2 regularization penalizes large weights by adding a term proportional to the square of the weights to the loss function, thereby encouraging the model to learn simpler, more robust patterns. These all help the model learn from rather simple and robust patterns. Specifically, L2 regularization was applied to the weights of the fully connected layer, which maps the LSTM outputs to the final forecast. This combination of L2 regularization and dropout provides a complementary approach to controlling overfitting, with dropout addressing neuron-level over-reliance and L2 regularization constraining the model's complexity by limiting excessively large weight magnitudes. Together, these techniques ensure a balance between model expressiveness and generalization to unseen data.

$$y = \mathbf{w}^T\mathbf{x} + b \tag{4.1}$$

The mean absolute error (MAE) is selected as the evaluation metric (loss function) for evaluating the models (both linear and deep ones) we used. The preliminary results of feeding our data into both linear (ARMA(3,2)) and deep models showed that we achieved a lower MAE with the proposed deep model with the MAE of 7.13 ppb for TVOC predictions and 9.47 ppm for CO_2 predictions.

4.4 Summary of contributions

- Comparing the statistical and deep neural network models' performance in forecasting IAQ metrics time series data

- Forecasting the IAQ metrics for 15 and 30 minutes, looking ahead with the MAE of 9.87 ppm for CO_2 and 7.13 ppb for TVOC forecasting (for 30 minutes ahead).

Chapter 5

Enhancing building ventilation through demand-driven strategies: balancing indoor air quality and ventilation rate

5.1 Abstract

Traditional ventilation systems in buildings often consume high levels of energy and can underperform in maintaining optimal indoor air quality (IAQ), sometimes failing to adequately keep pollutants like total volatile organic compounds (TVOC) and carbon dioxide (CO_2) within the recommended levels. This study introduces a novel approach to the operation of building ventilation systems, transitioning from traditional, schedule-based strategies to dynamic, demand-driven management to enhance energy efficiency and occupant well-being. By leveraging real-time data on IAQ, specifically concentrations of TVOC and CO_2, alongside occupancy patterns, our proposed method dynamically adjusts ventilation rates to maintain air quality within the recommended thresholds while optimizing energy usage. Through a four-month experimental analysis (two months of dynamic and two months of scheduled opera-

tion), we compared the performance of dynamic and scheduled ventilation operations in terms of pollutant levels and ventilation rates. Our findings reveal that dynamic operation reduces the average concentrations of TVOC and CO_2, with mean values of 128 ppb and 497 ppm under dynamic conditions, respectively, compared to 206 ppb and 544 ppm under scheduled operation. Furthermore, the dynamic approach achieved a notable decrease in ventilation rates during unoccupied periods, leading to overall energy savings without compromising IAQ during occupied times. Specifically, the study observed a reduction in total ventilation rates from 155 CFM under scheduled operation to 126 CFM under dynamic operation, underscoring the efficacy of the proposed method in enhancing both energy efficiency and indoor environmental quality.

5.2 Introduction & background

In response to the escalating energy costs of the 1970s in the United States [43], coupled with the renewed emphasis on energy efficiency sparked by the COP 21 conference in Paris in 2015 [89], there has been a significant transformation in construction practices. This evolution, targeting the substantial energy demands of the building sector—which accounts for 30-40% of total energy consumption—has prompted the development of buildings with improved air tightness. Notably, ventilation systems are among the leading contributors to energy consumption in buildings, responsible for approximately 48% of total building energy use [57]. Therefore, the primary goal of these design changes is to enhance thermal insulation and reduce energy consumption, a shift that is clearly reflected in the decreased air exchange rates in contemporary homes and office buildings [43, 89].

However, this increased air tightness has reduced opening areas, resulting in inadequate air exchange [43]. Moreover, harmful emissions from the insulation materials used in energy-efficient buildings and various other building materials and the accumulation of pollutants from occupants' activities have deteriorated indoor environmental quality. So, such design changes resulted in negative consequences to indoor air quality and an increase in building-related illnesses and sick building syndrome (SBS). These building-related symptoms were initially reported in the 1980s as the ventilation rate decreased [90]. Consequently, the need for effective ventilation strategies that can efficiently eliminate indoor air pollutants while maintaining optimal energy consumption has become increasingly critical. Previous research has shown that different approaches for optimal operations of the ventilation systems can yield annual energy savings of up to 30% [91, 92, 93].

Toward the goal of developing effective ventilation strategies, there have been several studies on smart buildings' ventilation systems and air purifiers to obtain the best IAQ and highest energy efficiency by adjusting the ventilation actions to the actual needs of the occupants [94]. This method of need-based control of ventilation actions is called demand-controlled ventilation (DCV), which should be highly energy efficient. The CO_2-based DCV has been one of the significant approaches followed by researchers in heating, ventilation, and air-conditioning (HVAC) in both academia and industry since decades ago [95]. In these studies, CO_2 was considered as the main effective proxy of IAQ and used to automate the ventilation systems.

In the CO_2-based DCV management strategy, to maintain decent indoor air quality and lower the energy consumption of the building HVAC system, the ventilation rates are modified based on the indoor CO_2 concentration [96]. Lu et al. [96] reviewed different studies focused on CO_2-based DCV systems and discussed the merits and demerits of different approaches toward these systems. They summarized these strategies into three: rule-based control, model-based control, and learning-based control methods. They discussed that, in the context of CO_2-based DCV management systems, most research focuses on demonstrating the effectiveness of model-based control strategies. However, these strategies are complex because ventilation systems are represented as nonlinear equations, and accurately estimating the control model's parameters and external inputs is challenging. As a result, implementing model-based control strategies in a field is challenging. Their other finding was that there exists a lack of field tests for advanced rule-based CO_2-based DCV strategies. They concluded that although this approach demonstrated its efficacy through simulation-based studies, there are limited field tests for this strategy in real building operations.

Furthermore, Lu et al. [96] observed that many studies on CO_2-based DCV strategies that considered occupancy have used CO_2 as the proxy of occupancy status, and there exists a gap for comparison between these studies and the ones that have used other sources for occupancy detection. Regarding the occupancy detection method, they also observed that ventilation reset control, which is a method that dynamically adjusts the outside air flow rate in a ventilation system, is also used to estimate real-time occupancy based on steady-state mass balance. Alternatively, transient equations and data-driven techniques could be used instead of steady-state equations for occupancy estimation. However, the effectiveness and advantages of these alternative approaches need to be proven through further testing and demonstration.

In their review paper, Lu et al. [96] concluded that the use of learning-based controllers in DCV systems is not yet fully mature. One major challenge mentioned for these controllers is the long training time required for reinforcement learning (RL) agents, which is a well-known issue not only in DCV systems but also in various other application domains. This issue also makes it challenging for real-time applications of these methods [97]. The other issue in many model-based and learning-based control strategies is that they only considered CO_2 for regulating the ventilation rates, and only a few of them have considered other metrics, such as temperature in their method. Their review [96] also underscored the significance of defining and adjusting various design and control parameters in rule-based DCV control strategies. These parameters include thresholds for outside air flow rates and the timing of system responses that can be obtained by the experts of the field as well. Further research is required to gain insights into the most optimal settings for these parameters in order to enhance ventilation system performance.

In some other studies, the ventilation systems' automation was implemented by simulating the occupants' ventilation behavior [98, 99]. Occupant ventilation behavior involves air exchange within a building, where occupants open or close windows or doors in response to changes in indoor environmental quality (IEQ) [89]. Incorporating the occupants' ventilation behavior into the building simulation programs can significantly decrease the discrepancies between simulated and actual energy use. Indeed, Andersen et al. [100] revealed that differences in occupants' ventilation behavior can result in differences in energy consumption of over 300%. These typical behaviors should be based on the quantification of real occupant behavior to improve the validity of the simulations' outcomes. So, one of the main incentives for researchers to do studies on modeling the occupants' ventilation behaviors is that it can improve the validity of the outcomes of energy use simulations. Another incentive of these studies is that obtaining an ideal indoor comfort level, both thermal comfort and IAQ comfort, wouldn't be achieved solely by relying on basic heat exchange principles, and it depends on how building occupants adjust their behaviors, such as heating, cooling, and opening windows or doors [99]. It's been shown that occupants who have the possibility to control their indoor environment have been found to be more satisfied and suffer from fewer building-related symptoms than occupants who occupy environments in which they have no control [101, 102]. So, one of the incentives of previous studies that focused on the occupants' ventilation behavior was that the occupants' comfort condition is as important as the existing guidelines [89] and should be considered in ventilation systems' automation [89].

Previous studies investigating occupant ventilation behaviors have examined various IEQ factors, including temperature and IAQ, as well as additional factors like time of day, outdoor climate conditions, and previous window states. These studies aimed to develop stochastic models for predicting the likelihood of occupants opening windows or doors [103, 104, 105, 106]. To achieve this, researchers observed different types of spaces over various time periods and collected relevant data. This data included both occupants' behaviors (considered as the ground truth for the target variable) and explanatory variables. These datasets were then used to train predictive models for occupants' ventilation behavior.

The research conducted by Rijal et al. [103] stands out as one of the studies that specifically delved into the influence of indoor thermal conditions on occupants' ventilation behavior. They conducted surveys in 15 office buildings across the UK, gathering data on indoor and outdoor temperatures over time. Notably, they excluded five air-conditioned (AC) buildings from their analysis, as their findings indicated that window openings were significantly less frequent in AC buildings compared to naturally ventilated ones. In these naturally ventilated buildings, windows played a crucial role in occupants' ventilation control. Utilizing the collected data, they employed logistic regression techniques to construct a stochastic model for predicting the likelihood of occupants opening windows. Their findings highlighted the potential benefits of improved building design, which correlated with enhanced comfort, reduced reliance on adaptive window behaviors, and a decrease in annual heating energy consumption. Moreover, their study suggested that an adaptive algorithm could better represent human-driven window control, providing a more accurate evaluation of human thermal comfort and building performance, including concerns like summer overheating and annual energy usage [103].

Herkel et al. [104] also developed a stochastic model to predict the likelihood of window openings in individual offices. Their model was not solely based on indoor and outdoor temperatures but also considered outdoor humidity. The data used for their study was gathered from 21 individual offices located within the Fraunhofer Institute's building in Freiburg, Germany, spanning from July 2002 to July 2003. Their findings revealed a robust correlation between the percentage of open windows and several factors, including the time of year, outdoor temperature, and patterns of building occupancy. They also observed that a significant portion of window openings occurred shortly after occupants arrived.

Several researchers have also considered the influence of IAQ factors on occupants' ventilation behaviors, going beyond the exclusive focus on thermal comfort aspects. For instance, Anderson et al. [105] conducted a comprehensive study in which they monitored indoor and outdoor environmental parameters alongside occupants' ventilation actions in 15 residential buildings located in Denmark. This investigation spanned from January to August and involved categorizing these buildings into four groups based on ventilation type (natural/mechanical) and ownership (owner-occupied/rented), aiming to uncover common trends in occupants' ventilation behaviors. Indoor measurements were collected every 10 minutes and included data on dry bulb temperature, relative humidity, illuminance, and CO_2 concentration. Occupants' ventilation behaviors were recorded, primarily noting whether windows were open or closed in all buildings and, in three dwellings, the actual window opening angle. Outdoor environmental factors, encompassing air temperature, relative humidity, wind speed, global solar radiation, and sunshine hours, were also assessed. Utilizing this comprehensive dataset, logistic regression models were developed to predict the likelihood of window opening and closing. The findings underscored the significance of two key variables: indoor CO_2 concentration, which served as an indicator of indoor air quality, emerged as the most influential factor for predicting window opening probability, while outdoor temperature played a pivotal role in determining the likelihood of window closure.

In another notable study conducted by Cali et al. [106], IEQ factors were meticulously measured in 60 apartments in Germany over a span of four years. These factors included temperature, relative humidity, CO_2 concentration, TVOC, ceiling light levels, infrared/visible light ratio, and the status of windows (open or closed), with data collected every minute. To assess the likelihood of window state changes, the researchers applied logistic regression to the extensive dataset they had collected. The analysis outcomes highlighted the primary factors influencing window openings, with more than 70% of the modeled windows being affected by the time of day and over 50% influenced by indoor CO_2 concentrations. Conversely, the decision to close windows was predominantly driven by the daily average outdoor temperature (for nearly 70% of the modeled windows) and, to a lesser extent, the time of day (for more than 50% of the modeled windows). Furthermore, examining data across different room types revealed that occupants' window-opening behaviors were shaped by various activities within the home, such as moisture-producing activities like showering and cooking, as well as by the presence of indoor pollutants.

Other than CO_2, TVOC is another indoor air pollutant that several studies have shown the importance of monitoring its levels in the indoor environment due to the effect it has on the occupants' well-being [107, 108]. Holos et al. [109] reviewed several studies regarding the impact of ventilation rates on the TVOC emission rates from the building materials and the TVOC concentration within indoor environments, especially in newly built or renovated buildings. Their review aimed to address two key questions: 1. Does an increase in ventilation rates affect the emission rates of TVOC in indoor environments, especially in newly constructed or renovated buildings? 2. Does the ventilation rate influence indoor TVOC concentrations, particularly in newly constructed or renovated buildings? Should these buildings consider initially boosting ventilation rates during the off-gassing phase to achieve acceptable IAQ? The findings from their review of relevant studies indicated that ventilation rate adjustments did not have a discernible impact on TVOC emission rates. Instead, the primary factor influencing the reduction of TVOC sources was time. These studies revealed that increasing ventilation rates did not significantly expedite the depletion of emission sources. Nevertheless, research consistently demonstrates that heightened ventilation rates play a crucial role in managing TVOC concentrations within indoor environments.

Research findings have demonstrated that in homes with very low infiltration rates, turning off ventilation led to a significant increase in indoor TVOC concentrations [110]. This underscores the crucial need to maintain minimum ventilation rates, even when there are no occupants present. Noguchi et al. [111] conducted a study measuring TVOC levels in two rooms with varying ventilation rates immediately after construction and three months later. Their results consistently showed that the room with higher ventilation rates maintained lower TVOC concentrations during both periods. Multiple studies [109] have established that newly constructed or renovated buildings emit indoor air pollutants at levels significantly higher (by two or more orders of magnitude) than established buildings. Enough ventilation must be provided in order to guarantee that TVOC concentrations stay within safe limits during occupied periods. Other than newly built buildings, there are various situations, depending on occupancy patterns and activities, where TVOC levels surpass safe thresholds, necessitating improved ventilation rates to maintain healthy indoor air quality. Consequently, these studies indicate the importance of adjusting ventilation rates during occupancy to control TVOC concentrations within safe and comfortable limits. During periods of non-occupancy, ventilation rates can be reduced significantly (though not completely turned off) as TVOC emission rates do not significantly accelerate with increased ventilation.

Previous methodologies for managing ventilation systems, while advantageous, exhibit several limitations. Firstly, there is a scarcity of models that concurrently account for the impact of multiple indoor air pollutants, such as CO_2 and TVOC, on occupants' ventilation behaviors. Secondly, existing approaches often fail to accommodate the diverse thermal sensations and IAQ preferences experienced by individuals in shared spaces, such as conference rooms, which complicates the prediction of ventilation behavior across occupants with varying comfort requirements. Thirdly, the primary mechanism by which occupants adjust ventilation rates, namely, the opening or closing of doors and windows, proves ineffective in environments devoid of windows or where maintaining an open door is impractical due to noise concerns. Therefore, there is a pressing need for innovative strategies to dynamically control ventilation systems that enhance IAQ and energy efficiency without relying on manual interventions by occupants.

In this experiment, to control the ventilation rates in occupied and unoccupied periods, we have used IoT devices to collect data on indoor air pollutant levels (CO_2 and TVOC) and occupancy numbers. The parameters contributing to determining the final ventilation rate are the current level of CO_2, TVOC, occupancy status, and future (15 minutes ahead) level of CO_2 and TVOC. To control the ventilation rates on occupied and unoccupied occasions, we defined a ventilation score that can be obtained using the five mentioned parameters, and then, based on different ranges of this score, we can switch the ventilation rate. Ventilation rates were determined in consultation with the facility management staff, who are experts in this domain. This experiment would overcome the mentioned limitations by considering CO_2 and TVOC at the same time to control the ventilation rate without any need for the occupants' interference. This field experiment, in which we integrate our method into the buildings' ventilation systems, would also overcome the scarcity of field tests and low integration of the theoretical methods into the buildings' systems.

5.3 Methodology

As previously mentioned, IAQ factors, particularly CO_2 and TVOC levels, have a significant impact on occupants' well-being and productivity. Conventional ventilation systems operate on predefined schedules, often without considering current occupancy and real-time IAQ metrics. Consequently, these systems may occasionally fall short of maintaining indoor air pollutant levels within the safe limits established by guidelines. Conversely, they may run at high rates even when the space is unoccupied, resulting in unnecessary energy waste. In this section, we outline our methodology for implementing a dynamic ventilation system in a conference room at the University of Virginia's Living Link Lab (L-LL).

The goal of this dynamic ventilation system is to save energy while simultaneously enhancing the comfort and health of occupants, allowing them to thrive within the indoor environment. Our method focuses on monitoring CO_2 and TVOC levels, as existing literature identifies these factors as significant contributors to indoor air quality and their potential impact on occupants' well-being and productivity [26, 42, 43]. This section provides an overview of the testbed (L-LL) and sensing system in the data collection subsection, followed by the data analysis and predictive models subsection and the experiment subsection, which goes over the methods we evaluated for dynamically operating the ventilation systems.

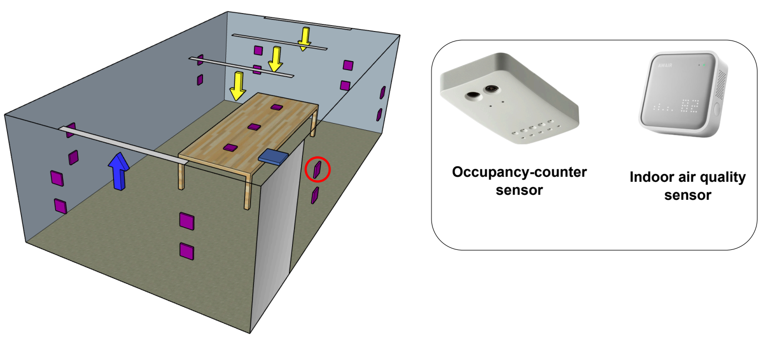


Figure 5.1: Left: sensors arrangement in the conference room. Purple cubes represent IAQ sensors, the blue cube represents the occupancy counter sensor, the red box is the selected location of the IAQ sensor for the data collection, and the yellow and blue arrows represent air inlet and outlet vents, respectively. Right: IAQ sensor and the occupancy counter sensor

5.3.1 Data Collection

The data for this study was collected in a conference room with a capacity of 20 occupants located in the L-LL at the University of Virginia campus. The L-LL is a 17,000 sqft open space office building and is occupied by roughly 150 residents, including faculty, research scientists, graduate students, and staff. The features and parameters of the sensors that are utilized in the L-LL for this study are summarized in Table 5.1. To count the number of occupants in the space, we have used the SenSource VIDIR series occupancy counter sensor. This sensor uses time-of-flight technology and counts the number of occupants in the area without invading the occupants' privacy. This device emits pulses of invisible infrared light, which bounces off surfaces beneath and back up to the sensor. The time taken for the pulses to bounce back allows the device to calculate how high up it is and how far away objects and people are. Importantly, these devices offer the flexibility to specify a desired detection height, a feature we found invaluable. It allowed us to adjust the sensor to disregard the door's movement as an occupant, ensuring accurate occupancy counts.

Sensor	Measurement	Value range	Error	Unit
SenSource VIDIR series	Occupancy count	-	3%	Person
Awair Element	CO_2 TVOC	[400, 5000] [0, 60000]	±75ppm ±10%	ppm ppb

Table 5.1: Characteristics of sensors used

To collect the IAQ-related parameters, we used Awair sensors, which can collect various parameters, including TVOC, CO_2, and light data. To optimize these sensor placements in the conference room, we initially installed nineteen sensors at various locations and heights (seating height of 3.94 and standing height of 5.91 ft.) as shown in Fig. 5.1). This pilot experiment was run during working hours for ten days (separate from the main study). Upon analyzing the collected data, we observed that all sensor readings were remarkably consistent, deviating by only 3-5% from each other. Utilizing both correlation analysis and random forest feature selection techniques, we identified one sensor (shown in the red circle in Fig. 5.1) as the most representative of the collective locations. This particular sensor's significance was further underscored by its strategic positioning, away from the direct influence of any ventilation vents, ensuring that its readings were not artificially impacted. Consequently, this location was selected for the experiment, balancing both statistical rigor and practical considerations in environmental monitoring.

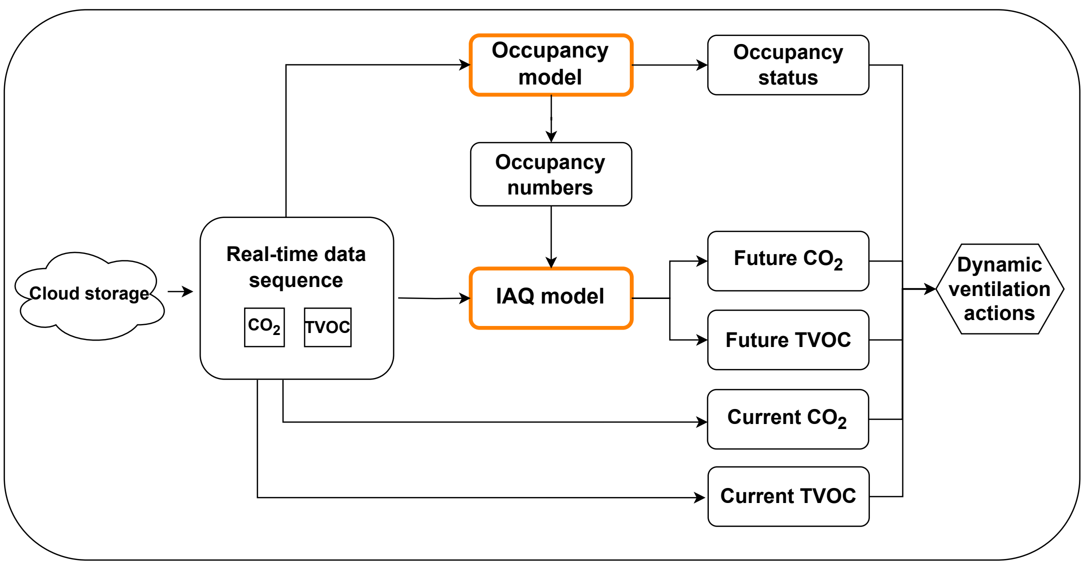


Figure 5.2: Proposed dynamic approach for ventilation system's operation

Figure 5.2 shows our proposed method for dynamically operating the ventilation systems. The features contributing to the final decision of the system are current CO_2 and TVOC levels, future CO_2 and TVOC levels (15 minutes ahead), and the current occupancy status. To forecast the future levels of these pollutants, we trained time series models, which use historical IAQ data and the number of occupants in the space as features. So, by having these pieces of information, we want to design an experiment in which we automate the ventilation system to: 1) make sure that the system is not operating at full capacity when it is not necessary (no occupants in the room) and 2) it would increase the ventilation rate when needed (high level of indoor air pollutants). The factors that have been used for controlling the ventilation systems include:

1. current CO_2 data : collected from the IAQ sensors

2. current TVOC data: collected from the IAQ sensors

3. Next 15-min CO_2 data: time series model that uses historical data of CO_2 levels and number of occupants in the space

4. Next 15-min TVOC data: time series model that uses historical data of TVOC levels and number of occupants in the space

5. Occupancy status: retrieved from the occupancy model (trained on the data collected by the occupancy counter sensor)

In the first phase (study 1) of the data collection, we collected 14 months (from 2022-06-30 to 2023-08-31) of TVOC, CO_2, and occupancy data to train the predictive models. In the second phase (studies 2 and 3), we collected data for four months to evaluate the performance of the dynamic operation of the ventilation systems and fine-tune the predictive models. During the first two months (from 2024-01-12 to 2024-03-12), the ventilation system was operated in dynamic mode, while in the second two months (from 2024-03-18 to 2024-05-18), it was operated in the default scheduled mode. Table 5.2 summarizes these studies and their usages.

5.3.2 Data Analysis & Predictive Models

We initially collected 14 months of data, including TVOC, CO_2, and the occupancy level. Utilizing this extensive historical dataset, we developed several predictive models aimed at different objectives, as depicted in Fig. 5.3. Initially, we trained a variety

Phase	Length of Study	Period	Usage	Ventilation Mode
Study 1	14 months	2022-06-30 to 2023-08-31	Training the predictive models	Default scheduled preset
Study 2	2 months	2024-01-11 to 2024-03-11	Implementing the dynamic operation of the ventilation systems and fine-tuning	Dynamic
Study 3	2 months	2024-03-18 to 2024-05-18	Evaluation and comparison of different ventilation modes' performance	Default scheduled preset

Table 5.2: Summary of study phases with their respective durations, periods, usages, and ventilation modes.

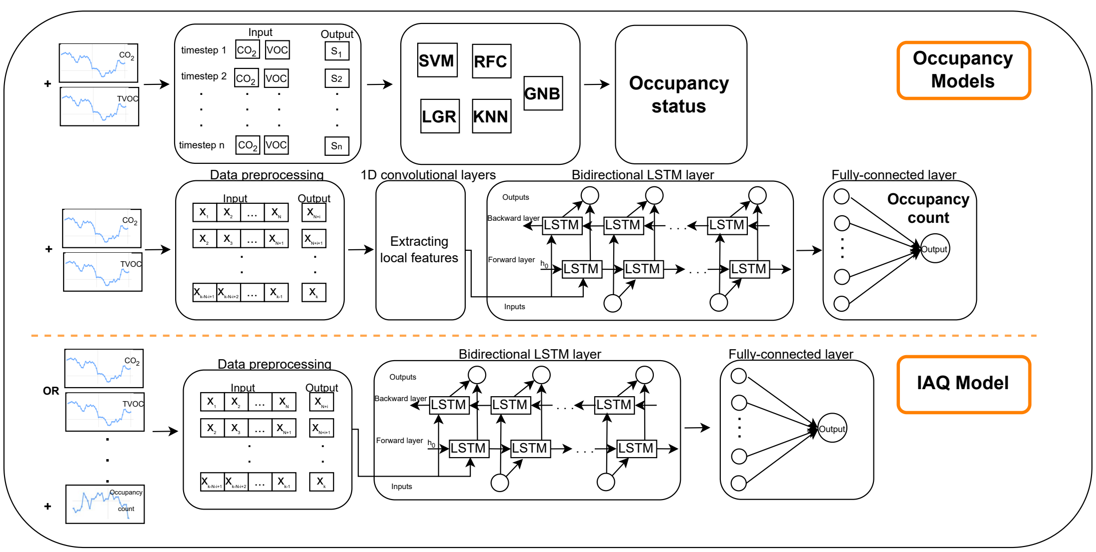


Figure 5.3: Predictive models for occupancy and indoor air quality parameters

of machine learning models to predict the binary occupancy status (0 indicating absence and 1 indicating the presence of at least one individual) using CO_2 and TVOC concentrations as predictive features. During this phase of model development, we encountered challenges with dataset imbalance, particularly due to the room being unoccupied more frequently than not over a 24-hour period. To address this issue, we narrowed our data selection window to the hours between 7 AM and 9 PM daily and applied the Synthetic Minority Over-sampling Technique (SMOTE) to achieve a balanced dataset.

Evaluation Metric	SVM (linear)	SVM (non_linear)	GNB	LGR	RFC	KNN	XGBoost
Accuracy	64	70	60	62	94	99	98
Precision	68	77	73	65	96	99	98
Recall	52	57	30	51	93	98	98

Table 5.3: ML models' performance for occupancy status prediction

To find the best predictive model, we trained and analyzed five different models: Support Vector Machine (SVM) with both linear and non-linear kernels, Gaussian Naive Bayes (GNB), Logistic Regression (LGR), Random Forest Classifier (RFC), K-Nearest Neighbors (k-NN), and eXtreme Gradient Boosting (XGBoost). According to the performance metrics summarized in Table 5.3, the k-NN model emerged as the most accurate, achieving a 99% accuracy rate. This result demonstrates how well the k-NN model can predict occupancy status based on CO_2 and TVOC levels with respect to precision and recall.

For the subsequent phase of predicting room occupancy levels, we employed both ML models and deep learning networks to ascertain their comparative effectiveness. Our analysis revealed that the Convolutional Bidirectional Long Short-Term Memory (CBLSTM) network significantly outperformed the traditional machine learning models previously utilized for binary occupancy status prediction. This approach begins with the raw CO_2 and TVOC sensor data being input into a convolutional network, which is structured around a convolutional layer paired with a max pooling layer. The convolutional layer efficiently processes the sequential data through a sliding filter window, extracting local features, while the max pooling layer is tasked with identifying and isolating the most discriminative of these features. This pooling layer highlights the essential characteristics and streamlines the model by reducing the overall number of features, consequently decreasing the number of parameters required [73]. This methodology enhances the model's efficiency and accuracy in predicting the precise number of occupants within a room.

Subsequently, the features extracted from the convolution network were input into a BLSTM network for further analysis. The LSTM network is distinguished by its proficiency in learning long-term dependencies within the data, a notable advancement over traditional Recurrent Neural Network (RNN) models [74]. Traditional RNNs, which rely on gradient backpropagation for training, often struggle with gradient vanishing and exploding problems. In contrast, the LSTM architecture incorporates specialized gates within its memory cells to regulate the flow of information, enabling it to preserve long-term dependencies effectively [74]. Our exploration of both unidirectional and bidirectional LSTM configurations revealed a critical insight that solely accounting for past temporal dependencies was insufficient for optimal model performance. The integration of future contextual information significantly enhanced the model's performance. Consequently, we opted for the BLSTM network as our final model due to its dual-direction processing capability, which systematically incorporates both past and future data contexts, offering a more comprehensive understanding of temporal dependencies.

Following that, the dense (fully connected) layers get the outputs from the BLSTM layer. The real occupancy count prediction is made in these layers, which also carry out the transition of the BLSTM output to the intended output shape. In order to translate features learned by the BLSTMs to the expected output, the fully connected layer mixes those features in nonlinear ways. The occupancy count in the conference room is derived from the patterns the model has discovered in the CO_2 and TVOC data. We utilized Mean Absolute Error (MAE) as the loss function, and the outcomes revealed the MAE of .04.

In the final step of the predictive modeling phase, we developed a deep neural network designed to forecast the levels of pollutants (CO_2 and TVOC) up to 30 minutes in advance. The architecture of this network is similar to the previously mentioned occupancy counter model, effectively capturing both the short-term and long-term characteristics inherent to the time series data of indoor air pollutants. The final layer (which is a fully connected one) forecasts the levels of pollutants (CO_2 and TVOC) up to 30 minutes in advance. This approach yielded an MAE of 9.47 ppm for 30 minutes ahead of CO_2 predictions and 7.13 ppb for 30 minutes ahead of TVOC forecasts, indicating the model's precision in predicting pollutant concentrations within indoor environments.

5.3.3 Experiment

The default scheduled operation of the HVAC systems sets the ventilation rate at 200 CFM from 6 AM to 7 PM and at 100 CFM from 7 PM to 6 AM without considering the real-time occupancy and IAQ parameters. In order to dynamically control the ventilation systems' operation, our study has evaluated two different approaches in detail: a non-linear approach and a linear approach. By using a linear approach, a linear ventilation score is created and used as the foundation for calculating the ventilation rate according to its calculated range. On the other hand, the non-linear approach uses a machine learning model, utilizing 14 months of historical data (from

2022-06-30 to 2023-08-31) to predict the optimal ventilation action. Results showed that the linear approach works better for its reliability and balanced response to CO_2 and TVOC levels, ensuring better occupant well-being and energy efficiency. In conditions of elevated CO_2 (above 1000 ppm), the linear model consistently increases the ventilation rates, while the non-linear one is unable to adequately boost ventilation rates during occupancy when CO_2 surpasses 1000 ppm. Additionally, the linear approach avoids the non-linear model's excessive sensitivity to TVOC levels, which can lead to unnecessary energy use. A detailed examination of each approach is given in the subsequent sections.

Parameter	Situation 1 (l_1)	Situation 2 (l_2)	Situation 3 (l_3)
Current CO_2 (s_1)	[0,1000)	[1000, 2500)	[2500, ∞)
Current TVOC (s_2)	[0,200)	[200, 500)	[500, ∞)
Occupancy status (s_3)	0	-	1
Future CO_2 (s_4)	[0,1000)	[1000, 2500)	[2500, ∞)
Future TVOC (s_5)	[0,200)	[200, 500)	[500, ∞)

Table 5.4: Parameters' situations for ventilation system automation

Linear approach

In the linear approach, we defined a linear ventilation score shown in equation 5.1, in which s_i corresponds to the values of contributing parameters and w_i is their associated weights. The contributing parameters and their associated weights are as follows:

s_1 represents the current CO_2 level and w_1 is its associated weight

s_2 represents the current TVOC level and w_2 is its associated weight

s_3 represents the current occupancy status and w_3 is its associated weight

s_4 represents the future CO_2 level and w_4 is its associated weight

s_5 represents the future TVOC level and w_5 is its associated weight

Table 5.4 shows different values that each s_i can take. Different ranges for the indoor air pollutants levels shown in Table 5.4 are selected based on the literature [43, 51, 42, 52] and their associated impact on the occupants' well-being. For l_1, l_2, and l_3 in Table 5.4, we have selected 0, 0.5, and 1, and the ventilation score can range from 0 to 1, accordingly.

$$S_{\text{Vent}} = \mathbf{w}^{\intercal}\mathbf{s} = w_1 s_1 + w_2 s_2 + w_3 s_3 + w_4 s_4 + w_5 s_5 \tag{5.1}$$

We have considered four ventilation ranges (conditions) to switch to them based on different values of the ventilation score (S_{Vent}). The ventilation rates were selected in consultation with the HVAC facility managers, who are experts in this domain. Here are the four ranges we have selected:

$0 \leq S_{\text{Vent}} \leq 0.25$: ventilation system would be set to 100 cfm (will be referred as v_1)

$0.25 < S_{\text{Vent}} \leq 0.5$: ventilation system would be set to 200 cfm (will be referred as v_2)

$0.5 < S_{\text{Vent}} \leq 0.75$: ventilation system would be set to 305 cfm (will be referred as v_3)

$0.75 < S_{\text{Vent}} \leq 1$: ventilation system would be set to 460 cfm (will be referred as v_4)

The first step toward getting w_i values is defining our desired conditions in terms of inequalities related to the ventilation score and then solving the inequality systems of our desired conditions to obtain the weights. In the following, we have defined our desired conditions, followed by their corresponding inequalities:

- Condition 1: When there is no occupant in the room and the current and future CO_2 and TVOC levels are in state l_1 or l_2 (please refer to the Table 5.4 for different states): We want the ventilation system to be in v_1 state (equivalent to 100 cfm).
  - CO_2 state: l_1, TVOC state: l_1, future CO_2 state: l_1, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_4 \leq 0.25$$

  - CO_2 state: l_1, TVOC state: l_1, future CO_2 state: l_2, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_3 \leq 0.25$$

  - CO_2 state: l_1, TVOC state: l_1, future CO_2 state: l_2, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_3 + \frac{1}{2}w_4 \leq 0.25$$

  - CO_2 state: l_1, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_2 \leq 0.25$$

  - CO_2 state: l_1, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_2 + \frac{1}{2}w_5 \leq 0.25$$

  - CO_2 state: l_1, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_2 + \frac{1}{2}w_4 \leq 0.25$$

– CO_2 state: l_1, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_2 + \frac{1}{2}w_4 + \frac{1}{2}w_5 \leq 0.25$$

– CO_2 state: l_2, TVOC state: l_1, future CO_2 state: l_1, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_1 \leq 0.25$$

– CO_2 state: l_2, TVOC state: l_1, future CO_2 state: l_1, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_5 \leq 0.25$$

– CO_2 state: l_2, TVOC state: l_1, future CO_2 state: l_2, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_4 \leq 0.25$$

– CO_2 state: l_2, TVOC state: l_1, future CO_2 state: l_2, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_4 + \frac{1}{2}w_5 \leq 0.25$$

– CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_2 \leq 0.25$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_5 \leq 0.25$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_1 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_4 \leq 0.25$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_2 :

$$0 \leq \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_4 + \frac{1}{2}w_5 \leq 0.25$$

- Condition 2: When there is at least one occupant in the room and the current and future CO_2, and TVOC levels are in state l_1: We want the ventilation system to be in v_2 state (equivalent to 200 cfm).

  - CO_2 state: l_1, TVOC state: l_1, future CO_2 state: l_1, future TVOC state: l_1 :

$$0.25 < \frac{1}{2}w_3 \leq 0.5$$

- Condition 3: When there is at least one occupant in the room and the current CO_2 and TVOC levels are in state l_2 and future CO_2, and TVOC levels are in state l_1 or l_2: We want the ventilation system to be in v_3 state (equivalent to 305 cfm).

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_1 :

$$0.5 < w_3 + \frac{1}{2}w_1 + \frac{1}{2}w_2 \leq 0.75$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_1, future TVOC state: l_2 :

$$0.5 < w_3 + \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_5 \leq 0.75$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_1 :

$$0.5 < w_3 + \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_4 \leq 0.75$$

- CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_2, future TVOC state: l_2 :

$$0.5 < w_3 + \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_4 + \frac{1}{2}w_5 \leq 0.75$$

- Condition 4: When there is at least one occupant in the room, and the current CO_2 and TVOC levels are in state l_2 and future CO_2, and TVOC levels are in state l_3: We want the ventilation system to be in v_4 state (equivalent to 460 cfm).

  - CO_2 state: l_2, TVOC state: l_2, future CO_2 state: l_3, future TVOC state: l_3 :

$$0.75 < w_3 + \frac{1}{2}w_1 + \frac{1}{2}w_2 + w_4 + w_5 \leq 1$$

These are the conditions that we want to be satisfied when we are automating the ventilation systems. So, we have solved this inequality system to see the solutions for w_is. Solving this inequality system resulted in the following solution:

- When $0 < w_1 \leq 0.25$, then:
  - $0.25 \cdot (1 - 4w_1) \leq w_2 \leq 0.5 \cdot (1 - 2w_1)$
  - $w_3 = 0.5$
  - $w_4 \leq 0.5 \cdot (1 - 2w_1 - 2w_2)$
- When $0.25 < w_1 < 0.5$, then:
  - $0 < w_2 \leq 0.5 \cdot (1 - 2w_1)$
  - $w_3 = 0.5$
  - $0 < w_4 < 0.5 \cdot (1 - 2w_1 - 2w_2)$

And in both of these cases, $w_5 = 1 - w_4 - w_3 - w_2 - w_1$.

In the next step to verify if different values of w_1 to w_5 within the obtained ranges in the solution would result in the same (consistent) ventilation actions (v_1-v_4), we have tested our approach in the historical data that we have. In this validation step, we have randomly set different values for w_1, w_2, and w_4 (within the defined ranges of the solution) to see if they would result in consistent ventilation actions. We evaluated the parameters and the ventilation ranges on historical data of 14 months, and the results showed that in 98% of the situations, different values of w_1, w_2, and w_4 resulted in consistent ventilation range selection. The other 2% showed two consequent ventilation ranges. So, the following values were selected for the weights to calculate the ventilation score and select the decent air exchange rate accordingly.

$w_1 = 0.15$

$w_2 = 0.25$

$w_3 = 0.5$

$w_4 = 0.05$

$w_5 = 0.05$

Machine learning (non-linear) approach

In addition to the linear decision-making approach, we used a machine learning approach to implement dynamic control of the ventilation systems for comparative analysis. In supervised machine learning, we needed to define features and targets. Similar to the previous approach, our features include current CO_2 and TVOC levels, occupancy status, and future CO_2 and TVOC levels. As for the feature variables, we leveraged historical data spanning 14 months, from June 2022 to August 2023. Notably, in this approach, the target variable is ventilation action, corresponding to the air exchange rate chosen to control pollutant levels effectively. In line with our previous approach, we considered the same four ventilation actions, each corresponding to different air exchange rates, denoted as v_1, v_2, v_3, and v_4. It's essential to note that during the data collection period, the ventilation systems operated according to a predefined schedule, with only v_1 and v_2 ventilation actions activated.

To determine the appropriate target variable for each of the feature sets, we adopted the same predefined boundaries (Table 5.4) as in the previous approach. For instance, when the room is unoccupied, and both the current and future CO_2 or TVOC levels fall within state l_1 or l_2 (refer to Table 5.4 for the various states), we assign the target variable as v_1 (defined as condition 1 in the previous approach). As another example, when there is at least one occupant in the room, and the current CO_2 and TVOC levels are in state l_2, while future CO_2 and TVOC levels are in state l_1 or l_2, we designate the target variable as v_3. For a comprehensive description of the different conditions used to define the target variables based on the feature range, please refer to the previous approach for dynamic ventilation system control (conditions 1-4).

After establishing both the feature and target variables, it became evident that the target variable exhibited a significant imbalance. The majority of instances corresponded to v_1, while v_4 was the minority class, accounting for less than 1% of the dataset. To address this issue, we made the decision to exclude instances where v_4 was selected as the target variable. Furthermore, we opted to exclude data collected during the hours of 8 PM to 5 AM, as this timeframe typically corresponds to periods with no occupants in the area, and as a result, during these times, v_1 was predominantly selected as the ventilation action. Subsequently, we applied SMOTE to rectify the imbalanced dataset. Before implementing SMOTE, we meticulously divided the data into separate sets for training, validation, and testing, ensuring no data leakage between these partitions. In our machine learning model selection, we considered four different models: LGR, RFC, Gradient Boosting Classifier (GBC), and SVM.

In the training phase of the machine learning models, we employed a comprehensive hyperparameter tuning approach using GridSearchCV, a method provided by the scikit-learn library. This technique systematically explores a range of hyperparameter configurations to identify the combination that yields the best model performance, as measured by predefined metrics such as accuracy, precision, recall, and F1-score. Our objective was to optimize the parameters of the classifier. As evident in Table 5.5, the RFC consistently outperforms the other models. In the following sections, we will delve into the training phase of the models, with a specific focus on RFC due to its superior performance and robustness in handling our dataset.

In the hyperparameter tuning phase, the GridSearchCV process was configured with the RFC as the estimator (for the RFC model). We defined a parameter grid that specified multiple values for various hyperparameters of interest, including the number of trees in the forest, the maximum depth of the trees, the minimum number of samples required to split an internal node, and the minimum number of samples required at each leaf node. This grid represented the exhaustive combination of hyperparameters to be evaluated. Also, to ensure the generalizability of the model and mitigate the risk of overfitting, we implemented a 5-fold cross-validation (CV) strategy. In this approach, the dataset was randomly divided into five equal parts. In each CV iteration, four parts were used for training the model, and the remaining part served as the validation set to assess model performance. This process was repeated five times, with each of the five parts used exactly once as the validation set. The cross-validation approach ensures that the assessment of hyperparameter effectiveness is robust and not overly dependent on a particular partitioning of the data.

Given the computationally intensive nature of performing grid search across multiple hyperparameter combinations and CV folds for several ML models, we leveraged parallel processing capabilities by setting n_jobs=-1. This configuration allowed the process to utilize all available CPU cores, significantly reducing the time required to complete the grid search. Upon completion of the grid search, GridSearchCV provided us with the best hyperparameter combination based on the average performance across all CV folds. The selected hyperparameters for RFC are n_estimators = 300, max_depth = 30, min_samples_split = 2, and min_samples_leaf = 1. We then evaluated this optimal model configuration on a separate test set to obtain unbiased estimates of its generalization performance. The metrics reported in Table

5.5 include precision, recall, F1-score, and support for each class, alongside overall accuracy, providing a comprehensive view of the model's predictive capabilities.

Model	Class	Precision	Recall	F1-score	Support
RFC	0	1	1	1	20754
	1	0.95	0.98	0.97	20650
	2	0.98	0.95	0.97	20699
	Accuracy	-	-	0.98	62103
	Macro avg	0.98	0.98	0.98	62103
	Weighted avg	0.98	0.98	0.98	62103
LGR	0	1	1	1	20754
	1	0.92	0.92	0.92	20650
	2	0.92	0.92	0.92	20699
	Accuracy	-	-	0.95	62103
	Macro avg	0.95	0.95	0.95	62103
	Weighted avg	0.95	0.95	0.95	62103
GBC	0	1	1	1	20754
	1	0.90	0.98	0.94	20650
	2	0.98	0.89	0.93	20699
	Accuracy	-	-	0.96	62103
	Macro avg	0.96	0.96	0.96	62103
	Weighted avg	0.96	0.96	0.96	62103
SVM	0	1	1	1	20754
	1	0.91	0.96	0.93	20650
	2	0.96	0.90	0.93	20699
	Accuracy	-	-	0.95	62103
	Macro avg	0.96	0.95	0.95	62103
	Weighted avg	0.96	0.95	0.95	62103

Table 5.5: Model Evaluation Metrics

The performance result of each model is shown in Table 5.5. As mentioned, the RFC outperformed the other models with an accuracy of 98%. Other performance metrics, including precision, recall, f1-score, and support, are reported as well. Precision measures the proportion of correct positive predictions for each class, i.e., among all instances predicted as a given class, how many are actually of that class. The RFC model shows perfect precision for class 0 (1.00), which is exceptional. Classes 1 and

2 also have high precision scores (0.91 and 0.96, respectively), indicating a high level of reliability in the predictions for these classes. Recall (also known as sensitivity) measures the ability of the model to detect all actual instances of a given class. RFC showed perfect recall for class 0 (1.00) and very high recall for classes 1 and 2 (0.96 and 0.90, respectively). This means that this model can identify all instances of class 0 and performs well on classes 1 and 2, too.

F1 score is the harmonic mean of precision and recall, providing a single metric to assess the balance between them. RFC achieves perfect balance for class 0 (1.00) and has very high F1 scores for classes 1 and 2 (0.93 for both), indicating a well-balanced performance between precision and recall. In our project, the F1 score emerges as the paramount metric for evaluating the performance of our model, given its critical role in harmonizing precision and recall, which are two metrics of equal importance to our objectives. Precision (the proportion of true positive results in all positive predictions) is critical because it reflects the model's ability to correctly identify instances that require adjustment in the ventilation system. High precision means that when the model predicts the need for changes in ventilation settings (e.g., to improve air quality or reduce energy consumption), those predictions are likely correct, minimizing unnecessary adjustments that could waste energy or negatively impact air quality.

On the other hand, recall (the proportion of true positive results in all actual positives) is equally important because it measures the model's ability to detect all instances that truly need intervention. High recall ensures that the model identifies as many real situations as possible where adjustments to the ventilation system are needed, thus maintaining indoor air quality without compromising occupant comfort. The F1 score harmonizes these two metrics, ensuring that improvements in one do not disproportionately come at the expense of the other. In this project, optimizing for the F1-score means we aim for a ventilation system that responds accurately and comprehensively to energy efficiency and air quality needs without leaning too much towards avoiding false alarms (precision) or missing real events (recall).

Support values reported in Table 5.5 refer to the number of actual occurrences of each class in the dataset. The distribution of instances across the classes is nearly uniform, which helps in evaluating the model's performance more evenly. Accuracy gives the proportion of all predictions (across all classes) that were correct. RFC's overall accuracy is 0.95 (or 95%), which is outstanding, especially considering that this is an average across three classes. Macro Avg and weighted Avg for precision, recall, and F1-score provide aggregated performance metrics. A macro average computes the metric independently for each class and then takes the average (hence treating all classes equally), while a weighted average takes into account the support of each class. Both averages, being around 0.95 and 0.96, indicate consistently high performance across all classes, with or without considering the number of instances per class.

The next step of our framework involves a real-world experiment of dynamically managing the ventilation system in a conference room by using either the linear or non-linear method. As it is a real-world experiment in a conference room where occupants have meetings, it is imperative to ensure the chosen approach not only enhances the ventilation rate in response to suboptimal air quality, as determined by predefined criteria, but also avoids inefficiencies that could lead to unnecessary energy consumption. To identify the most effective method that satisfies these criteria, we conducted a preliminary trial. This trial involved simulating adjustments to the ventilation rates without actual implementation, aiming solely at assessing the performance of both the linear and non-linear models.

For the execution of this pilot study, we utilized real-time data on CO_2 and TVOC, applying our predictive model to forecast their values 15 minutes into the future. Although the forecasting models can accurately predict IAQ parameters up to 30 minutes in advance, consultations with experts led us to select 15 minutes as the optimal timeframe for making necessary adjustments to the ventilation rates. In addition to these IAQ-related parameters, by using the occupancy models we've trained, we obtained the occupancy status and level which are essential for both methodologies to determine the most suitable ventilation rate. Beyond applying these approaches to data derived from real-world conditions, we created datasets for hypothetical extreme scenarios not encountered during the actual data collection period. Although the likelihood of such extreme cases is minimal, based on our historical observations, it was crucial to test the resilience and adaptability of our methods under these hypothetical conditions to ensure their robustness in the face of potential real-world challenges.

For running the experiment using any of the approaches, we retrieved 10 historical data points (with 5-minute intervals, which is 50 minutes of data in total) of indoor air pollutants (CO_2 and TVOC) and the occupancy level. Then, we used these data as the features of our trained forecasting models and predicted 15 minutes ahead of the indoor air pollutants levels. Then, in the linear approach, based on these predictions, the obtained parameters (w_1 - w_5), and the occupancy status of the space, we calculated the S_{Vent} and selected the ventilation state (v_1, v_2, v_3, and v_4) based on the ventilation score. On the other hand, for the non-linear approach, the retrieved information fitted into the RFC model to predict the best ventilation rate.

The results of the first trial revealed that:

- The non-linear approach demonstrates limitations in its capacity to elevate ventilation rates to levels above the baseline during periods of occupancy (identified as v_2), particularly when CO_2 concentrations exceed 1000 ppm, corresponding to the l_2 and l_3 categories as outlined in the Table 5.5. Evaluations conducted under hypothetical scenarios indicated that this approach fails to adjust the ventilation to higher levels when CO_2 levels fall within the l_2 and l_3 ranges, and the TVOC concentration is within the l_1 range (below 200 ppb). This suggests that the non-linear system is predominantly responsive to TVOC levels. The underlying issue is due to the rarity of CO_2 levels exceeding 1000 ppm in this specific setting, according to the historical data used for training the model. Consequently, the machine learning model, being trained on historical data, lacks the necessary conditioning for scenarios where CO_2 levels exceed 1000 ppm. Despite the infrequency of such occurrences in historical records, it is critical for the model to possess the capability to escalate ventilation rates appropriately in response to elevated CO_2 levels, considering the potential adverse effects on occupants, especially during prolonged exposure periods. On the other hand, the linear approach consistently proves to be effective in activating higher ventilation rates when CO_2 concentrations are within the l_2 or l_3 ranges. This reliability makes the linear method more favorable, as it better ensures the occupants' well-being.

- Despite its shortcomings in responding to elevated CO_2 levels, the non-linear model exhibits an overcautious approach to TVOC levels during occupancy in the conference room. Specifically, it identifies the threshold for TVOC at approximately 185 ppb and elevates the ventilation rate to a level beyond the baseline (which is v_2) whenever TVOC concentrations exceed this threshold. While this sensitivity is advantageous for occupant health, it potentially leads to unnecessary energy use. Conversely, the linear model adopts a more measured approach to TVOC management, initiating an increase in ventilation rate to above the baseline only when current or predicted TVOC levels reach or exceed 200 ppb. For instance, should the current TVOC concentration be 194 ppb with a forecasted rise to 201 ppb, the linear model would adjust the ventilation to a level higher than the baseline. The linear approach, guided by the carefully determined thresholds in Table 5.5, is preferred for its balance between ensuring occupant well-being and optimizing energy use. The linear method's strategy underscores the unnecessary extent of the non-linear model's proactive measures, which, while well-intentioned, result in energy inefficiency.

- During periods of non-occupancy, when the current and future concentrations of selected indoor air pollutants are categorized within the l_1 and l_2 ranges, both methodologies demonstrated proficiency in adjusting the ventilation system to the v_1 setting. The performance of both approaches was equivalent and satisfactory under these conditions.

- During periods of non-occupancy, when both current and future concentrations of the selected indoor air pollutants reach critical levels (both in the l_3 range), the linear approach proactively adjusts the ventilation system to v_2, a level higher than the baseline (v_1) set for non-occupied periods. In contrast, the non-linear method maintains the ventilation at the minimum level, v_1. In such situations, we prefer the performance of the linear method. This preference is based on the anticipation that the room might be occupied shortly, and immediate occupants would be subjected to significantly high levels of indoor air pollutants until the ventilation system can improve the air quality. Hence, the linear method's preemptive adjustment to a higher ventilation rate is crucial for ensuring the well-being of future occupants by reducing potential exposure to harmful pollutants.

Drawing from the outcomes observed during the trial experiment, we have elected to employ the linear method in the main experiment. This decision is driven by our objective to simultaneously improve the occupants' well-being and optimize energy consumption. To conduct the experiment and dynamically adjust the ventilation rates in response to occupancy status, as well as current and future levels of CO_2 and TVOC, we partnered with the UVA facility management. This collaboration granted us control over the operation of the conference room's ventilation system. Real-time data and the predefined linear approach were used to change the ventilation rates at 15-minute intervals, which is a frequency determined through consultations with the facility management team.

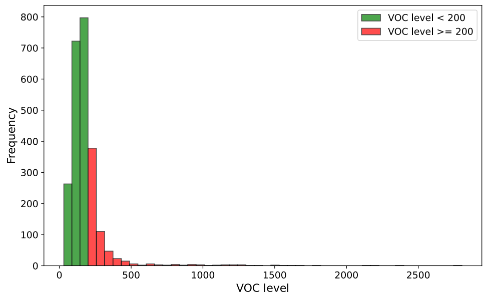
As shown in Fig. 5.2, after one month of running the experiment and collecting new IAQ and occupancy data under the dynamic operations of the ventilation system, we proceeded to fine-tune both the IAQ forecasting and occupancy prediction models. The reason for this step is that these models were initially trained on the data that was collected during the scheduled operation of the ventilation systems, and we want to make sure that the models are still accurate during the dynamic operation of the ventilation system. The rest of the experiment was implemented using the updated models. To maintain model performance and adapt to any further changes in the data patterns, a final round of fine-tuning was conducted upon the completion of the entire experiment (two months).

5.4 Results

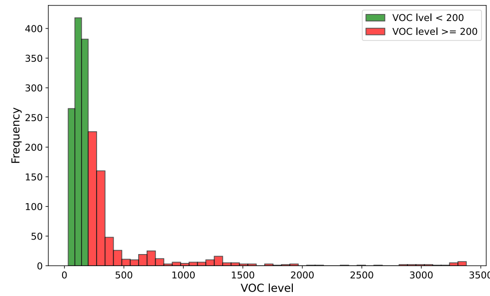
In this section, we present findings on the frequency at which the selected indoor air pollutants (CO_2 and TVOC) exceed established thresholds, alongside ventilation rates observed under various conditions of HVAC operation, both dynamic and static. By comparing these metrics, we aim to deepen the understanding of the effectiveness of our proposed method for the dynamic management of ventilation systems. Beyond aggregate results spanning the entire experimental period (two months), we showcased the variations in TVOC and CO_2 levels and the ventilation rates, comparing instances of dynamic ventilation system operation to periods of scheduled operation. This detailed comparison during selected days of the experiment underscores the potential benefits of our approach.

In the initial phase of our analysis, the main goal is to assess various parameters derived from IAQ data collected under two different operational modes of the ventilation systems: scheduled and dynamic. The main goal of this evaluation is the assessment of ventilation rates, TVOC, and CO_2 levels. The focus on IAQ is crucial as it pertains to occupants' exposure to indoor air pollutants and the subsequent potential adverse health impacts. Furthermore, a comparison of the ventilation rates under the two operating situations (dynamic and planned) of the ventilation systems will make clear how energy-efficient our suggested method is.

Fig. 5.4 presents histograms illustrating the distribution of TVOC during both dynamic and scheduled operations of HVAC systems, while Fig. 5.5 depicts the corresponding data for CO_2 concentrations. Throughout the observation period, the mean TVOC concentration was found to be 128 ppb under dynamic operation and 206 ppb during scheduled operation. Similarly, CO_2 concentrations averaged at 497 ppm and

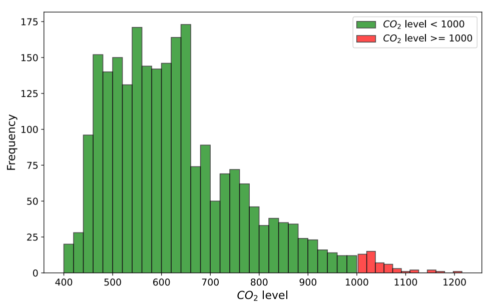


(a) During dynamic operation of the ventilation systems

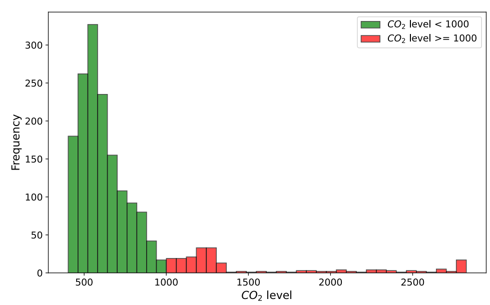


(b) During scheduled operation of the ventilation systems

Figure 5.4: Histogram of TVOC levels during dynamic and scheduled operation of the HVAC systems



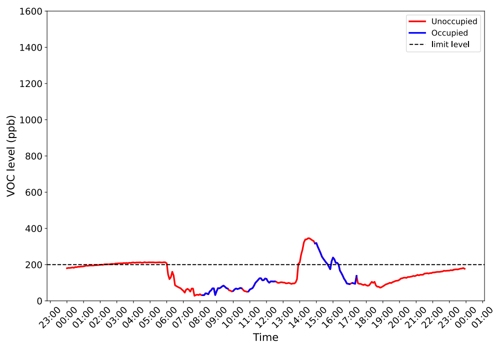
(a) During dynamic operation of the ventilation systems



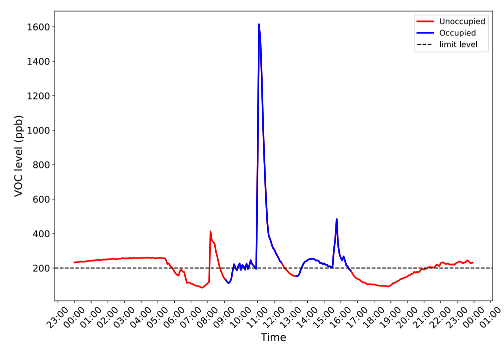
(b) During scheduled operation of the ventilation systems

Figure 5.5: Histogram of CO_2 levels during dynamic and scheduled operation of the HVAC systems

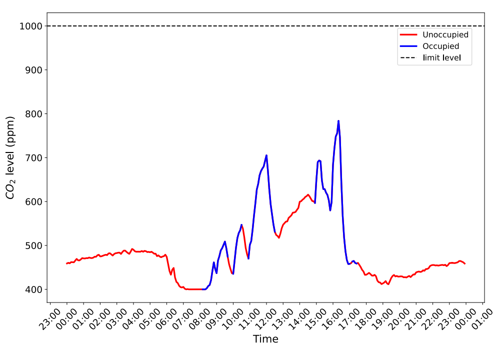
544 ppm for dynamic and scheduled operations, respectively. Additionally, Table 5.6 shows the average values of TVOC and CO_2 concentrations alongside the ventilation rates during periods of occupancy, non-occupancy, and the entire observation period for both ventilation strategies. It is observed that the mean concentrations of both TVOC and CO_2 are lower under dynamic ventilation compared to scheduled operation across all periods examined. This lower indoor air pollutant levels during occupied periods would lead to enhancements in the well-being of the occupants.



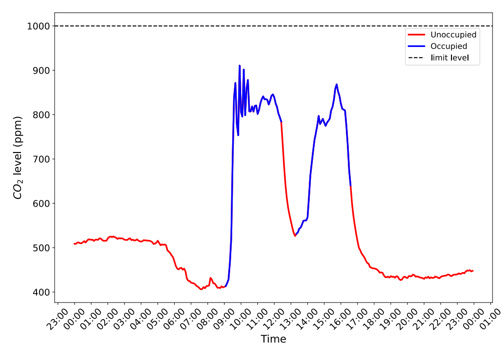
(a) TVOC level of a sample day with dynamic operations of the ventilation system



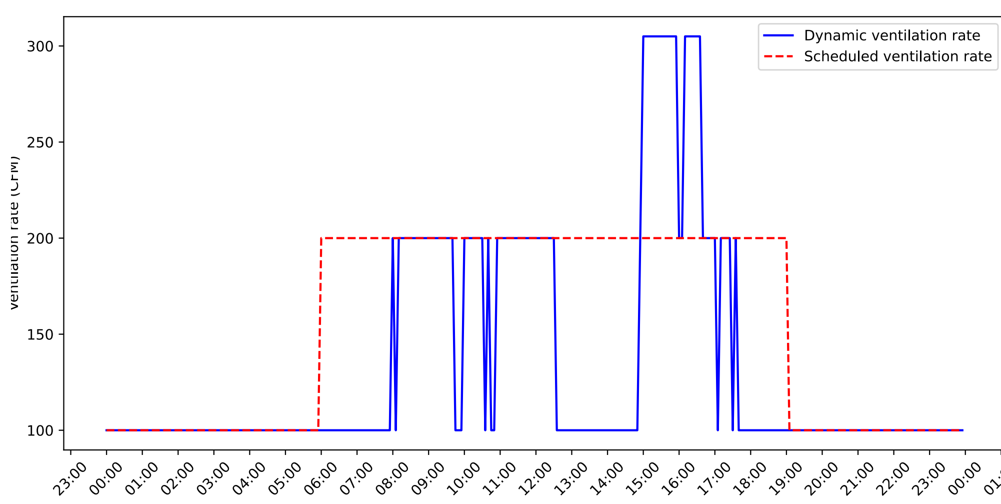
(b) TVOC level of a sample day with scheduled operations of the ventilation system



(c) CO_2 level of a sample day with dynamic operations of the ventilation system

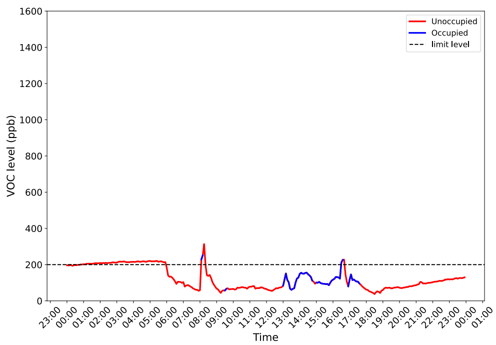


(d) CO_2 level of a sample day with scheduled operations of the ventilation system

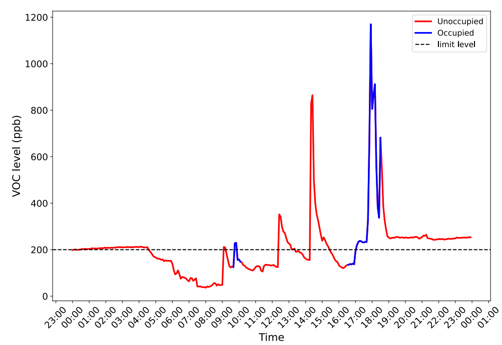


(e) Dynamic and scheduled ventilation rates

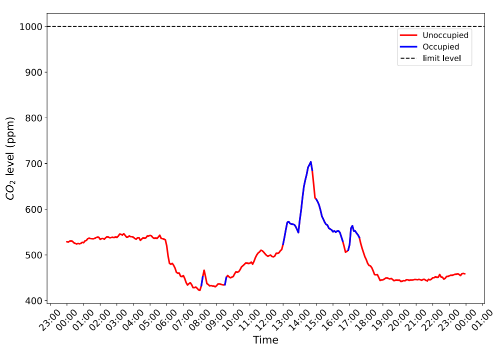
Figure 5.6: IAQ parameters' level of sample days with different ventilation system operations



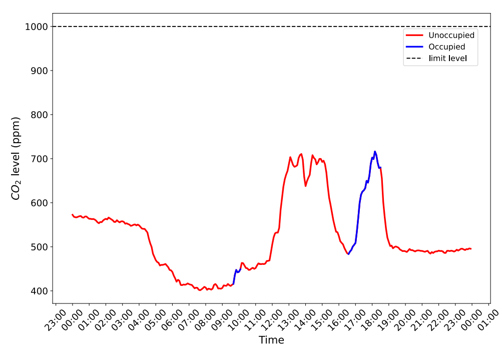
(a) VOC level of a sample day with dynamic operations of the ventilation system



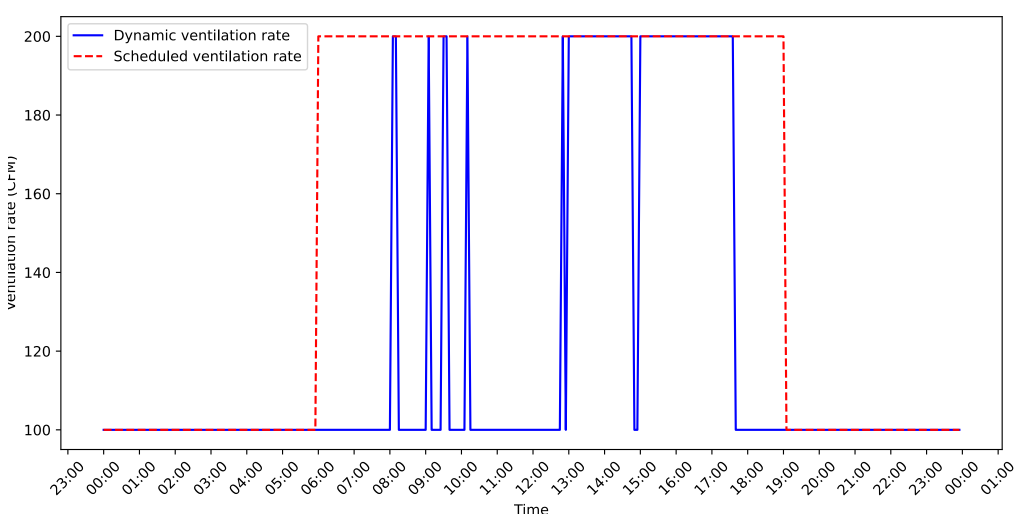
(b) TVOC level of a sample day with scheduled operations of the ventilation system



(c) CO_2 level of a sample day with dynamic operations of the ventilation system



(d) CO_2 level of a sample day with scheduled operations of the ventilation system



(e) Dynamic and scheduled ventilation rates

Figure 5.7: IAQ parameters' level of sample days with different ventilation system operations

	Occupied		Unoccupied		Overall	
	Dynamic	Scheduled	Dynamic	Scheduled	Dynamic	Scheduled
TVOC (ppb)	184	290	116	192	128	206
CO_2 (ppm)	628	722	468	514	497	544
Ventilation rate (CFM)	219	191	106	147	126	155

Table 5.6: Mean value of IAQ parameters and ventilation rate during scheduled and dynamic operation of ventilation systems

The data in Table 5.6 reveal that the average ventilation rate is higher during periods of occupancy under dynamic operation of the ventilation systems compared to their operation on a scheduled basis. Conversely, the lower average ventilation rates observed during unoccupied periods under dynamic operation contribute to a reduction in the overall average ventilation rates for the entire period of dynamic operation in comparison to scheduled operation (126 vs. 155 CFM). Thus, by optimizing ventilation rates—increasing them during occupied periods (219 vs. 191 CFM) and decreasing them during unoccupied periods (106 vs. 147 CFM), dynamic operation of the ventilation systems achieves a reduction in total ventilation rates (126 vs. 155 CFM), ultimately leading to energy savings. So, the results show that not only did the dynamic operation of the ventilation systems lead to lower ventilation rates, but it also resulted in lower TVOC and CO_2 levels and fewer instances of these pollutants exceeding their limit threshold.

To better understand the factors driving increases in ventilation rates to V4, we analyzed instances where these higher rates were dynamically selected. By calculating the sum of the parameter scores for all V4 instances, we found that occupancy was the most significant factor, followed by current TVOC levels, future TVOC levels, current CO_2 levels, and future CO_2 levels, in that order. Notably, the higher contribution of TVOC levels compared to CO_2 levels indicates that TVOC concentrations more frequently exceed acceptable thresholds in the conference room. This finding aligns with the results presented in Chapter 1, where we demonstrated that TVOC levels more often surpassed recommended limits, even when CO_2 concentrations remained within acceptable ranges. These results underscore the importance of monitoring and dynamically addressing TVOC levels to maintain healthy indoor air quality, particularly in spaces with fluctuating occupancy.

To further evaluate the energy efficiency of the dynamic mode, we replaced instances where the system selected 305 CFM with 460 CFM. This adjustment was made to explore if the dynamic mode would still be the more energy-efficient one. Despite the increased ventilation rate, analysis shows that dynamic operation would still lead to energy savings. In particular, if the system had run at 460 CFM in those periods when it chose 305 CFM— likely resulting in further reductions in air pollutants— the average ventilation rate in dynamic mode would be 127 CFM, still below the scheduled mode average of 155 CFM. That means that even if we had chosen higher ventilation rates during the dynamic mode, it would, overall, be more energy-efficient given the energy savings during unoccupied periods.

We further investigated the sensitivity of the calculations to the inclusion of TVOC and/or future CO_2 and TVOC levels. We systematically excluded these input parameters and recalculated ventilation rates, comparing the three different periods of overall, occupied, and unoccupied. Tables 5.7, 5.8, and 5.9 present the average ventilation rates for the three periods under the various input combinations. As we can see in Table 5.7, during the overall period of the dynamic operation, decreasing the input parameters has actually resulted in higher ventilation rates, except in the condition that we only considered current CO_2 as the driving factor of the ventilation actions.

As we can see in Tables 5.7, 5.8, and 5.9, when we only consider current CO_2 as the input parameter, the mean ventilation rate is 105 CFM, with the breakdown of mean ventilation rate of 116 CFM during occupied periods and 103 CFM during unoccupied periods (shown in Tables 5.8 and 5.9). This is while if we only consider TVOC as the input parameter, the overall ventilation rate increases to 164 CFM, which is higher than the condition we considered all the input parameters. As shown in Tables 5.8 and 5.9, only considering TVOC as the input parameter would increase the ventilation rate in unoccupied periods too, which might be unnecessary as no occupant is being impacted. These findings confirm our findings in Chapter 1 that it happens more frequently for TVOC to surpass its limit range compared to CO_2 exceeding its limit range.

Overall period							
Current CO_2	**Current TVOC**	**Current occupancy**	**Future CO_2**	**Future TVOC**	**Mean ventilation rate (CFM)**	**IAQ**	**Energy**
✓	✓	✓	✓	✓	126	✓	✓
✓	✓	✓			135	✓	✗
✓		✓			132	✗	✗
	✓	✓			140	✓	✗
✓					105	✗	✓
	✓				146	✓	✗

Table 5.7: Overview of input parameter combinations and their impact on mean ventilation rates and IAQ detection during dynamic ventilation operation. The first row serves as the baseline, using all parameters for decision-making. The ”Energy” column indicates whether a combination uses less energy (✓) or more energy (✗) compared to the baseline, as reflected in the mean ventilation rate (CFM). The ”IAQ” column assesses the ability of each combination to detect conditions where either CO_2 or TVOC exceeds acceptable limits, with ✓indicating effective detection and ✗ indicating failure to detect all exceedance scenarios.

Although interpreting the IAQ conditions for each combination requires experimental validation, we assessed the impact of these combinations by examining the scenarios in which they lead to increased ventilation rates. It should be noted that all these combinations for dynamic operation of the ventilation system resulted in better performance regarding IAQ and energy compared to the scheduled operation. In Table 5.7, to compare their performance, the combination in which all the parameters are considered (Which is the method we used for running the experiment) is used as the baseline.

As mentioned, in this table, the evaluation criteria are categorized into two aspects: Energy and IAQ columns. The meaning of the ✓and ✗ in each of these columns is as follows:

1. **Energy:**
   - A checkmark (✓) in this column indicates that the combination uses less energy compared to the baseline method. Energy consumption is represented by the mean ventilation rate (CFM), so a lower value signifies better energy efficiency.
   - An X mark (✗) signifies that the combination uses more energy than the baseline, indicated by a higher mean ventilation rate.
2. **IAQ:**
   - A checkmark (✓) indicates that the combination successfully detects conditions where either CO_2 or TVOC levels exceed their acceptable limits. This ensures adequate indoor air quality control.
   - An X mark (✗) means the combination fails to identify all cases where either CO_2 or TVOC exceeds the acceptable range. For example:
     - For example, if only CO_2 is considered, the system may miss cases where TVOC levels are too high. However, if only TVOC is considered, it may still account for both TVOC and CO_2 exceedances due to their overlap in detection capabilities.

As we can see in Table 5.7, including occupancy as an input variable to TVOC, would decrease the mean ventilation rate, with an increase in the ventilation rate during the occupied periods and a decrease in the ventilation rate during the unoccupied periods. It shows that considering occupancy would appropriately shift the ventilation rates from unoccupied periods to occupied ones to improve the occupants' comfort. As we can see, by adding the current CO_2 level to TVOC and current occupancy status, we reassigned importance weights between these three parameters (compared to TVOC and occupancy), which reduced the focus on TVOC and resulted in lower ventilation rate. As we can see in Table 5.7, the lowest ventilation rate after the condition in which we only considered current CO_2 level is when we considered all the parameters with their different associated weights. This shows that when we put the focus on one parameter, it increases the ventilation rate unless it is CO_2, which happens less frequently when it exceeds its limit range. Although this analysis provides better insight into the inclusion of different parameters, we need to keep in mind that ventilation rates and IAQ parameters impact each other, and we need to further establish this sensitivity analysis by running experiments for each different input parameter combination.

Occupied period					
Current CO_2	**Current TVOC**	**Current occupancy**	**Future CO_2**	**Future TVOC**	**Mean ventilation rate (CFM)**
✓	✓	✓	✓	✓	262
✓	✓	✓			328
✓		✓			324
	✓	✓			363
✓					116
	✓				164

Table 5.8: Overview of input parameters and mean ventilation rates during the occupied period of dynamic ventilation operation.

Unoccupied period					
Current CO_2	**Current TVOC**	**Current occupancy**	**Future CO_2**	**Future TVOC**	**Mean ventilation rate (CFM)**
✓	✓	✓	✓	✓	103
✓	✓	✓			103
✓		✓			100
	✓	✓			103
✓					103
	✓				142

Table 5.9: Overview of input parameters and mean ventilation rates during the unoccupied period of dynamic ventilation operation.

Overall period									
Current CO_2	Current TVOC	Current occupancy	Future CO_2	Future TVOC	Mean ventilation rate (CFM)	v1=100 CFM	v2=200 CFM	v3=305 CFM	v4=460 CFM
✓	✓	✓	✓	✓	126	83%	10%	5.0%	2.0%
✓					105	96%	4.0%	-	0.26%
	✓				146	64%	32%	-	4.0%
Occupied period									
Current CO_2	Current TVOC	Current occupancy	Future CO_2	Future TVOC	Mean ventilation rate (CFM)	v1=100 CFM	v2=200 CFM	v3=305 CFM	v4=460 CFM
✓	✓	✓	✓	✓	262	-	56%	34%	10%
✓					116	88%	11%	-	2.0%
	✓				164	62%	27%	-	10%
Unoccupied period									
Current CO_2	Current TVOC	Current occupancy	Future CO_2	Future TVOC	Mean ventilation rate (CFM)	v1=100 CFM	v2=200 CFM	v3=305 CFM	v4=460 CFM
✓	✓	✓	✓	✓	103	97&	3.0%	-	-
✓					103	97&	3.0%	-	0.4%
	✓				142	65%	33%	-	3.0%

Table 5.10: Overview of the ventilation rates frequencies in percentage during different combinations of input parameters

Table 5.10 provides a detailed breakdown of ventilation rate distributions across three combinations of input parameters: (1) all input parameters, (2) only the current CO_2 level, and (3) only the current TVOC level. When TVOC levels alone are considered, the ventilation rate is elevated to 460 CFM during unoccupied periods 3% of the time. Additionally, 33% of the time, the ventilation rate is set to 200 CFM, which is significantly higher compared to the case where all input parameters are considered, where 200 CFM is used only 3% of the time, and the ventilation rate defaults to 100 CFM the rest of the time. Given that the space is unoccupied during these periods, the frequent elevation of ventilation rates when only TVOC is used as the input parameter results in unnecessary energy consumption.

In contrast, when the current CO_2 level is the sole parameter, the distribution of ventilation rates during unoccupied periods closely resembles the distribution when all parameters are considered, with only 0.4% of cases operating at 460 CFM. During occupied periods, however, relying solely on the current CO_2 or TVOC levels results in the system operating at 100 CFM, which is the minimum recommended ventilation rate accepted in our case, for 88% and 62% of the time, respectively. This highlights an inadequacy for occupied periods, where higher ventilation rates are typically required for acceptable IAQ. It is important to note that the dataset used for this sensitivity analysis, comprising CO_2, TVOC, and occupancy data, was collected while the ventilation system operated in scheduled mode.

Fig. 5.6 shows the TVOC, CO_2, and ventilation rates of sample days with dynamic and scheduled operation of the ventilation systems. We selected the same day of the week to keep most of the meetings constant in both of the sample days. On this day, there were meetings with a high number of occupants that resulted in a higher mean value of the ventilation rates during the occupied period when the ventilation systems were operating dynamically compared to their scheduled operation. The mean value of the ventilation rates during occupied periods is 224 CFM and 200 CFM during dynamic and scheduled ventilation operations, respectively. These values are 100 and 137 CFM in unoccupied periods during dynamic and scheduled ventilation operations, respectively. Overall, the mean level of the ventilation rates is lower during the whole period of dynamic ventilation systems' operation compared with the scheduled operation of the ventilation systems (134 CFM vs. 155 CFM) on this specific day of the week.

Fig. 5.7 displays the TVOC, CO_2 concentrations, and ventilation rates for two additional sample days. On these days, all meetings had fewer attendees, eliminating the necessity to elevate ventilation rates during periods of occupancy. Consequently, for both dynamic and scheduled operations of the ventilation systems, the average ventilation rate during occupancy was maintained at 200 CFM. The mean value of the ventilation rates during unoccupied periods is 103 CFM and 143 CFM during dynamic and scheduled ventilation operations, respectively. As a result, the dynamic operation that reduced ventilation rates during unoccupied times led to lower overall ventilation rates for the entire day compared to the scheduled operation (122 CFM vs. 155 CFM).

5.5 Conclusions and Discussion

In conclusion, the research presented in this paper offers significant insights into the benefits of dynamically operating ventilation systems over conventional, schedule-based operations. Our comprehensive analysis reveals that the dynamic management of ventilation systems improves IAQ by effectively maintaining lower concentrations of key pollutants such as TVOC and CO_2 and enhances occupant well-being by ensuring air quality remains within the recommended thresholds more consistently. Importantly, this approach demonstrates a clear advantage in optimizing energy usage, achieving energy savings without compromising indoor environmental quality.

The findings from our experimental analysis, including the distribution of TVOC and CO_2 concentrations alongside ventilation rates under both dynamic and scheduled operations, underscore the efficacy of the proposed method. Specifically, we observed lower mean concentrations of TVOC and CO_2 under dynamic operation across all periods examined (occupied and unoccupied periods), highlighting the method's potential to reduce occupants' exposure to harmful indoor air pollutants.

Our approach's ability to dynamically adjust ventilation rates based on real-time occupancy and pollution levels translates into a more energy-efficient operation, as evidenced by the reduced overall average ventilation rates during dynamic operation compared to the scheduled operation. This operational efficiency is achieved without sacrificing the indoor environmental quality, as indicated by the fewer instances of pollutant levels exceeding their limit thresholds and overall lower concentration of indoor air pollutants under dynamic management.

The research presented here convincingly argues for a shift towards dynamic ventilation systems in buildings to achieve a dual goal of energy efficiency and enhanced IAQ. Such systems represent a forward-looking solution to maintaining healthy indoor environments, while also meeting the urgent need for energy efficiency in building operations. As buildings evolve to become more intelligent and responsive, adopting dynamic ventilation strategies offers a promising pathway for achieving these objectives, holding the potential to make buildings more sustainable and occupant-friendly through impacts on building design, operation, and management practices.

5.6 Summary of contributions

The contributions of this section are as follows:

- Proposing a model-based and rule-based dynamic operation for the ventilation systems and comparing their performance
- Operating the ventilation system dynamically and comparing its performance regarding IAQ metrics and ventilation rate with the traditional, scheduled-based operation
- Preparing a data frame to fine-tune the predictive models for the situation of the dynamic operation of the ventilation system

Chapter 6

Introducing an IAQ index based on the expected performance loss according to the indoor environmental metrics' levels

6.1 Abstract

The airtight building designs introduced in the 1970s, aimed at reducing energy costs, have significantly impacted indoor air quality (IAQ), contributing to health issues such as sick building syndrome (SBS). High concentrations of indoor pollutants, including carbon dioxide (CO_2), total volatile organic compounds (TVOC), and particulate matter ($PM_{2.5}$), alongside poor thermal comfort, have reduced occupant health and productivity. This chapter introduces a novel performance loss index that integrates these IAQ metrics, temperature, humidity, and occupancy, to quantify the impact on occupant well-being and productivity. In this chapter, we have defined two different indices to evaluate and compare the IAQ and thermal comfort during dynamic and scheduled operation of the ventilation system. We included temperature and humidity to define an index to evaluate the impact of thermal comfort on the

occupants. We call this index the thermal comfort index. For defining the other index to evaluate the impact of IAQ on the occupants, we included CO_2 and TVOC levels. We call this index IAQ index. We collected the required data to acquire the values of these indices by using the data collected from a conference room over a four-month period, which included two months of dynamic HVAC operation and two months of scheduled operation. Results showed that dynamic ventilation reduced both indices, with the thermal comfort index decreasing from 0.25 To 0.18 and the IAQ index from 0.53 to 0.12, demonstrating improvements in IAQ and thermal comfort under dynamic operation. These results showed the efficacy of dynamic ventilation systems in reducing pollutant levels and improving occupant productivity.

6.2 Introduction & background

The rise in energy costs during the 1970s prompted a shift in construction methods across the United States, with buildings being designed to be more energy-efficient and airtight. This resulted in a decrease in the amount of air exchanged in homes and office buildings, with the average air exchange rate for homes dropping from around one air change per hour (ACH) to about 0.5 ACH during this period [112, 113]. Persily [114] outlines the initial ASHRAE 62 standard, introduced in 1973, and its numerous subsequent versions (such as ASHRAE 62.1, which pertains to commercial buildings), highlighting how our knowledge about the correlation between ventilation rates and acceptable indoor air quality has evolved over time. He also outlined that Similar to the history of residential ventilation, the requirements for commercial ventilation were also reduced in the early 1980s, primarily to conserve energy [114].

When buildings are made more air-tight, it can limit the exchange of air between the indoor and outdoor environments, which can lead to an accumulation of indoor air pollutants, such as CO_2 and TVOC. This can result in reduced indoor air quality, leading to various negative health effects, including headaches, fatigue, and respiratory problems. For example, the emergence of building-related illnesses and sick building syndrome (SBS) can be traced back to the 1980s, when ventilation rates began to decline, as documented by Riesenberg et al. in 1986, [115]. Also, Fisk et al. [116] highlighted that this trend has resulted in substantial yearly expenses and decreased productivity due to health issues associated with indoor environments.

Several indoor air pollutants have major impacts on the occupants' health and productivity. As mentioned in chapter 2, based on the literature, CO_2 and TVOC are the main indoor air pollutants that can adversely impact the occupants' productivity and health at their high concentration. Other than these two pollutants, $PM_{2.5}$ can also adversely affect the occupants' health. Most of the studies [117, 118] have focused on the effect of ambient $PM_{2.5}$ on the occupants' health, but in recent years, several studies [119] found indoor $PM_{2.5}$ to have comparable to or even greater adverse health effects compared with ambient $PM_{2.5}$. There are different sources of indoor $PM_{2.5}$ [120] such as smoking, cooking, human activities, and the ambient $PM_{2.5}$ that enter residences through ventilation of buildings and air filtration [121]. Literature showed that $PM_{2.5}$ influences health more than it influences productivity. The health issues caused by $PM_{2.5}$ can range from respiratory problems, such as coughing and wheezing, to more serious conditions, such as asthma, heart disease, and stroke [121, 120].

Other than the indoor air quality parameters, temperature also has an impact on the occupants' performance. Temperature effects on the occupants can be of different types, such as occupants' comfort, energy level, and performance. When the temperature of an indoor space is too high or too low, it can be difficult for the occupants to focus well on their work, leading to a decrease in occupants' productivity level [122]. Akimoto et al. [123] conducted different experiments to investigate the relationship between thermal comfort and productivity in a task-conditioned office and found that deviation from thermal neutral conditions led to productivity loss. Also, Lan et al. [124] investigated the impact of air temperature on office workers' well-being and productivity through subjective ratings. They found that high temperatures can negatively affect productivity. On the other side, discomfort and distraction brought on by low temperatures might reduce productivity, too. According to the authors, keeping a suitable temperature range of 20–26°C can improve employees' productivity and well-being. All these studies highlighted the importance of considering the impacts of indoor temperature on office workers' productivity and providing an environment with suitable thermal comfort to increase the occupants' productivity.

Due to the mentioned importance of IAQ and thermal comfort for the occupants, it should be noted that the efficient ventilation rates should not reduce the occupants' comfort (regarding both IAQ and thermal comfort) while reducing energy costs. Regarding the importance of considering indoor environmental quality, several studies defined IAQ indices using different approaches. Some of the different approaches that were proposed in this regard are [125]: (1) One index per single pollutant, (2) simple aggregation, and (3) aggregation according to sources of pollutants and/or types of pollutants. The first method involves creating a dimensionless index by comparing the concentration of a pollutant to a predetermined reference value that represents a usual level of health risk. However, other risk indicators, such as odor or irritation thresholds, can also be used as the reference value. A ratio greater than one in this method, meaning that the pollutant's concentration exceeds its reference value, denotes a possible concern about indoor air quality. As an example of this approach, Teichman et al. [126] defined separate indices for different IAQ metrics such as CO_2, TVOC, and $PM_{2.5}$. They plotted the concentrations of pollutants against a reference concentration and determined whether the resulting data fell within or outside of specified ranges, indicating whether the building was meeting or failing to meet IAQ targets.

The second method, an individual index is created by combining the indices separately derived for each pollutant. The aggregation process in this method involves adding individual indices, taking the maximum index, or using other integration techniques. As an example of this method, the work by Cohas et al. [127] can be mentioned in which they chose the highest index among the indices computed for each individual pollutant to produce a global index that represents the worst-case scenario. As other examples of this method, Gadeau et al. [128] Castanet et al. [129] a simple algebraic sum to derive one index. Also, getting an average was used by Chiang et al. [18] to combine indexes for different contaminants. And finally, in the third approach, pollutants' indices are aggregated according to their categories or origins. For example, Quad et al. [130] arranged the pollutants into four different groups based on their sources. The different groups they considered were human sources, cooking, particles, and pollutants from the potential sources of gaseous pollutants.

Although previous methods in defining indices for IAQ metrics were useful for incorporating them in ventilation management, they have several limitations, such as considering metrics separately, not including the number of occupants, and only considering one reference (limit) value for each pollutant or indoor environmental quality metric. In these methods, the number of occupants was not directly added to the definition of indices, and they only contributed to the indices by impacting the indoor environmental quality metrics. To address these limitations, in this section, we considered three main indoor pollutants that affect occupants' health and performance, which are CO_2, TVOC, and $PM_{2.5}$, and two metrics that contribute to the occupants' thermal comfort, which temperature and humidity. We also directly input the number of occupants as a variable to the proposed index.

6.3 Methodology and results

In this section, we will use the data we collected in the conference room in the previous section to derive an index indicating the impact of IAQ and thermal comfort on the occupants' well-being and productivity. The period of data collection, as explained in chapter 5 in detail, is four months, including two months of data with the dynamic operation of the ventilation systems and two months with the scheduled operation of the ventilation systems. The collected data consists of CO_2, TVOC, temperature, and humidity level. The first two were used as the metrics of the IAQ, and temperature and humidity were used as the proxy of thermal comfort. In the first step, we defined individual indexes for each of these metrics by considering the performance loss that is associated with them at their different levels. For this goal, we have not considered one reference (limit) level for them; instead, we used previous experiments that derived different levels of performance loss associated with different levels of each of these metrics.

Research from [43, 42, 131] defined the loss of productivity due to CO_2 concentrations, which was built on a baseline of 600 ppm, with 20% loss at 1000 ppm, 50% loss at 1500 ppm, and 100% loss at 3000 ppm. These data were trend-fitted in Microsoft Excel to develop the following loss function for CO_2:

$$y_{\mathrm{CO_2}} = (-1.02\times10^{-8}{x_{\mathrm{CO_2}}}^3+4.27\times10^{-5}{x_{\mathrm{CO_2}}}^2+1.67\times10^{-3}x_{\mathrm{CO_2}}-14.17)\times10^{-2} \quad (6.1)$$

The effect of TVOC on productivity was defined using [43, 132, 52], and a curve was built on a baseline of 200 ppb, 50% loss at 600 ppb, 75% loss at 1000 ppb, and 100% loss at 2000 ppb. These data were similarly trend-fitted and produced the following loss function for TVOC:

$$y_{\mathrm{TVOC}} = (2.85\times10^{-8}{x_{TVOC}}^3-1.29\times10^{-4}{x_{\mathrm{TVOC}}}^2+0.213x_{\mathrm{TVOC}}-37.798)\times10^{-2} \quad (6.2)$$

The temperature was also included in the model due to its effect on occupant comfort. The loss function for temperature was based on a "goal zone" between 20 and 22.5 degrees C, with significant losses mounting below 15.5 degrees C and above 26.8 degrees C. The curve had the following equation:

$$y_{temp} = (-9.21\times10^{-3}{x_{temp}}^4+7.79\times10^{-1}{x_{temp}}^3-22.69{x_{temp}}^2+264.32x_{temp}-960)\times10^{-2} \quad (6.3)$$

For our model, we also reviewed different studies on the effect of humidity on the occupants' comfort, health, productivity, and risk of infection. A study [133] done in Swedish dwellings observed low relative humidity (below 15%) over two weeks during the winter season. They found possible correlations between low humidity and specific health symptoms. Another study [134] evaluated the impact of different humidity levels on students' learning performance. Their results showed that the students' performance decreased in relative humidity of 20% compared with 40%. Also, they observed significantly more fatigue and distraction at a relative humidity of

20% compared with 40%. Furthermore, their study concluded high relative humidity (above 60%) decreased the students' productivity.

There are several other studies that showed that occupants would experience lower stress levels and higher productivity at a relative humidity of above 40% and below 60% [135, 136, 137, 134]. In a study [136], the authors recommended relative humidity of at least 40% as they observed a significant association between low RH and elevated prevalence of upper airway symptoms, most clearly at relative humidity levels of below 38%. Another research also observed that office workers who were exposed to relative humidity levels between 30% and 60% were more likely to experience 25% less stress, which was measured by lower heart rate variability [137].

Based on these studies and guidelines, we defined specific thresholds for productivity loss. We considered the ideal humidity range of 40%-60% with no significant adverse impact on occupants' productivity and health (0% productivity loss). However, as humidity levels deviate from this optimal range, discomfort increases, leading to a decrease in productivity. For our model, we assumed that humidity levels between 30%-40% and 60%-70% result in a 20% reduction in productivity. For more extreme conditions in which the humidity level is below 30% or above 70%, we considered 50% of productivity loss. For the very extreme cases of humidity levels below 20% or above 80%, which represents a situation where the environment becomes completely unsuitable for normal functioning due to extreme discomfort and possible health risks like mold growth or respiratory distress, we assigned 100% of productivity loss. Based on these values, the fitted third-degree polynomial that describes the productivity loss as a function of humidity had the following equation:

$$y_{humid} = (2.60 \times 10^{-4} {x_{humid}}^3 + 5.03 \times 10^{-2} {x_{humid}}^2 - 6.98 x_{humid} + 190.15) \tag{6.4}$$

In each of these four equations, x is the metric's level at the time, and y is the expected performance loss associated with the metric. Then, for the purpose of occupants' efficiencies, we defined the variable f for each of the metrics derived by subtracting one from each of the metrics' y ($f_i = (1 - y_i)$). Then, to get the total efficiency based on each f_i, I used weighted geometric mean with the following equation:

$$\bar{f} = \left(\prod_{i=1}^{n} f_i^{w_i} \right)^{1/\sum_{i=1}^{n} w_i} = \exp\left(\frac{\sum_{i=1}^{n} w_i \ln f_i}{\sum_{i=1}^{n} w_i} \right) \tag{6.5}$$

where $\bar{f}$ is the total efficiency we get from considering all the metrics, and w_i is the weights associated with each of them. For the weights of CO_2, TVOC, $PM_{2.5}$, and temperature, which are the metrics that we considered in this section, we used the suggestion by [138]. In their work, they found the associated weights for different environmental metrics by asking several experts, and their results indicated 0.209 for IAQ-related metrics (CO_2 and TVOC) weight and 0.208 for thermal comfort-related metrics (temperature and humidity).

The final total loss can be calculated using equation 6.6, in which N is the number of occupants that we will get from chapter 3.

$$Loss_{\text{total}} = (1 - \bar{f}) \times N \tag{6.6}$$

The final goal of this section is to obtain an index that is an aggregation of indices defined for five different metrics. So, based on the value of each metric, individual indices will be calculated and aggregated using the geometric mean method. The final index indicates a quantifiable metric for evaluating the total performance loss in the space (conference room) based on the concentration of the IAQ metrics and temperature and humidity in each time step.

The data used in this section comes from the four-month dataset gathered in Chapter 5. This dataset was chosen because it includes two months during which the ventilation system was operated dynamically and two months when it was operated on a scheduled basis. This comprehensive dataset provides the opportunity to analyze how the dynamic operation of the ventilation system impacted occupant performance loss. To achieve this, we considered two sets of metrics for the performance loss index. The first set includes IAQ-related metrics: CO_2 and TVOC. The second set is the ones related to the occupants' thermal comfort, which are temperature and humidity.

After preprocessing the data and normalizing the indices for each environmental metric (ensuring that their respective weights determine the extent of their contribution to the final index), we derived the IAQ and thermal comfort indices. The results of the analysis showed that the IAQ index for the scheduled period of ventilation operation was 0.53 and for the dynamic operation of the ventilation system was 0.12. The thermal comfort index, which reflects the impact of temperature and humidity levels on occupant performance loss, was 0.25 under scheduled ventilation and 0.18 under dynamic ventilation. These results showed that the dynamic operation of the ventilation system has not only improved the IAQ-related metrics (TVOC and CO_2) but also has improved the thermal comfort-related metrics (temperature and humidity). So, although temperature and and humidity were not directly considered in our method for dynamically controlling the ventilation system, the dynamic operation still led to improvements in these metrics.

6.4 Summary of contributions

In summary, the contribution of this section is as follows:

- Quantifying the impact of dynamic ventilation system operation on reducing the overall performance loss of occupants
- Integrating occupant count into the performance loss calculation to provide a more accurate assessment of IEQ impacts
- Considering more than a single IAQ metric for defining the performance loss index
- Utilizing multiple reference values to derive individual indices for each metric, selected based on their specific impact on occupant well-being
- Employing geometric mean for index aggregation, ensuring a more robust and meaningful representation under extreme conditions of indoor pollutants and temperature

Chapter 7

Conclusion

7.1 Thesis overview

This thesis explores the importance of dynamic and comprehensive indoor air quality (IAQ) management in building ventilation systems, with a focus on optimizing occupant health, well-being, and energy efficiency. The chapters systematically develop and analyze advanced methods for monitoring and managing IAQ, combining traditional statistical techniques with cutting-edge machine learning models and introducing novel indices to assess indoor environmental quality based on productivity loss of the occupants exposed to poor indoor environmental quality (IEQ).

In Chapter 2, the limitations of conventional HVAC systems that rely mainly on carbon dioxide (CO_2) as the only indicator of IAQ is established. The need for a broader approach, including other pollutants such as volatile organic compounds (VOCs), is emphasized to improve IAQ in building environments. This chapter presents a four-month longitudinal study that evaluates the effectiveness of VOCs and CO_2 as IAQ indicators. This study examines several indoor spaces, such as conference rooms and single-occupancy offices, identifying instances of poor IAQ despite acceptable CO_2 levels and evaluates the occupants' exposure to these instances. The chapter emphasizes the necessity for a dynamic ventilation system that considers a broader spectrum of indoor air pollutants, beyond solely CO_2, to enhance occupant health protection.

In Chapter 3, we implemented occupancy detection by using IEQ metrics as the features. By analyzing environmental factors like CO_2, VOC, light, temperature, and humidity, this chapter evaluates which combination of these factors most effectively detects occupancy in real time. Machine learning models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) are applied, demonstrating that TVOC is a valuable indicator for detecting occupancy in some cases. This chapter's results recommend integrating TVOC data into the occupancy detection models to improve their performance and, as a result, improve the HVAC management systems.

Chapter 4 focuses on forecasting IAQ metrics using statistical models (ARMA) and deep learning models (LSTM and CNN-LSTM). The research shows that while linear models like ARMA can provide reasonable accuracy, deep learning models better capture the complex and nonlinear behavior of indoor air pollutants. A proposed CNN-LSTM framework, combining convolutional layers with bidirectional LSTM networks, improves forecasting accuracy, enabling prediction of pollutant levels up to 30 minutes in advance. This chapter's results provide us with future values of the indoor air pollutants that can be helpful information for the dynamic operation of the ventilation systems.

Finally, Chapter 5 details a four-month field study comparing traditional, schedule-based ventilation systems with dynamic, demand-driven ventilation operations. The findings show that the dynamic approach of the ventilation system operation is effective in reducing the concentration of pollutants while achieving energy savings. The dynamic approach demonstrated potential for reducing CO_2 and TVOC levels (improving IAQ) while lowering ventilation rates during unoccupied periods (enhancing energy efficiency) by utilizing real-time IAQ and occupancy data.

In Chapter 6, the thesis introduces the IAQ productivity loss index and thermal comfort productivity loss index. IAQ performance loss index incorporates two IAQ metrics, which are CO_2 and VOC, while the thermal comfort performance loss index incorporates temperature and humidity. Both indices have considered occupancy level too to quantify the impact of IAQ and thermal comfort on occupant productivity and well-being. To apply these indices, we used the data that we collected in the previous chapter, which is a four-month study in a conference room. This dataset includes two months of dynamic operation and two months of scheduled-based operation of the ventilation systems. The results showed that using the dynamic approach for operating the ventilation system has improved both indices, indicating less performance loss during this period.

7.2 Recommendations and future directions

This thesis provides valuable insights into IAQ and ventilation control; however, several areas could be further improved in future research to refine the conclusions and address limitations observed throughout the study. In the following, I will discuss the areas that can be improved in future research.

Expansion of Monitoring Duration and Environments: The dynamic approach proposed in this thesis for operating the ventilation systems has been tested over four months within a university setting. To generalize the findings more broadly, future work should aim to extend the data collection period to cover an entire year, capturing seasonal variations that may impact IAQ and building energy consumption. For example, during winter, occupants' clothes might impact the TVOC level, leading to the need for a higher ventilation rate during occupied periods. Additionally, expanding the test environments to include a wider variety of building types (e.g., residential, commercial, and industrial) would provide a more comprehensive understanding of how dynamic ventilation systems perform across different contexts.

It is worth mentioning that the computational cost of dynamic ventilation operation primarily arises during the training phase of deep learning models, where significant resources are required to process data and optimize the model parameters. However, once the model is trained, the inference phase used to make real-time decisions is computationally lightweight and well-suited for deployment in real-world applications. Given the energy savings and improved indoor air quality achievable with a dynamic operation, expanding its implementation across more spaces justifies the initial computational investment. The benefits in terms of efficiency, sustainability, and occupant comfort outweigh the relatively minimal ongoing computational costs associated with inference.

Inclusion of Additional IAQ Metrics: Although this thesis includes a wide range of environmental factors to define the index that evaluates the performance of dynamic ventilation systems, the dynamic operation itself only controls CO_2 and TVOC levels. Incorporating a broader range of environmental factors in the dynamic control of the ventilation system can improve the results and efficacy of the ventilation systems.

Occupant Feedback Incorporation: For defining the performance loss index, this study uses literature for defining the loss associated with different values of the indoor environmental metrics. However, a comprehensive feedback system can be employed in areas of interest, such as conference rooms or single-occupancy offices, to improve the results. This feedback loop would not only enhance the comfort and satisfaction of the occupants but also contribute to the optimized performance of the ventilation systems regarding energy consumption. These feedback systems can be more helpful in single-occupancy offices, where the primary resident is a constant person.

7.3 Publications

This research project resulted in a few journal and conference publications. A comprehensive list is presented below.

- Varnosfaderani, Mahsa Pahlavikhah, Arsalan Heydarian, and Farrokh Jazizadeh. ”A longitudinal study of IAQ metrics and the efficacy of default HVAC ventilation.” Building and Environment 254 (2024): 111353.
  Chapter 2 covers the material in this publication [108].

- Varnosfaderani, Mahsa Pahlavikhah, Arsalan Heydarian, and Farrokh Jazizadeh. ”Using Statistical Models to Detect Occupancy in Buildings through Monitoring VOC, CO_2, and Other Environmental Factors.” Computing in Civil Engineering 2021. 2021. 705-712.
  Chapter 3 covers the material in this publication [55].

Bibliography

[1] World Health Organization et al. *WHO guidelines for indoor air quality: selected pollutants*. World Health Organization. Regional Office for Europe, 2010.

[2] TG Theodosiou and KT Ordoumpozanis. "Energy, comfort and indoor air quality in nursery and elementary school buildings in the cold climatic zone of Greece". In: *Energy and Buildings* 40.12 (2008), pp. 2207–2214.

[3] Zufeng Pei et al. "Comparative study on the indoor environment quality of green office buildings in China with a long-term field measurement and investigation". In: *Building and Environment* 84 (2015), pp. 80–88.

[4] Krystallia K Kalimeri et al. "Indoor air quality investigation of the school environment and estimated health risks: two-season measurements in primary schools in Kozani, Greece". In: *Atmospheric Pollution Research* 7.6 (2016), pp. 1128–1142.

[5] Concettina Marino, Antonino Nucara, and Matilde Pietrafesa. "Thermal comfort in indoor environment: Effect of the solar radiation on the radiant temperature asymmetry". In: *Solar Energy* 144 (2017), pp. 295–309.

[6] Neil E Klepeis et al. "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants". In: *Journal of Exposure Science & Environmental Epidemiology* 11.3 (2001), pp. 231–252.

[7] Joseph G Allen and Linsey C Marr. "Recognizing and controlling airborne transmission of SARS-CoV-2 in indoor environments". In: *Indoor air* 30.4 (2020), p. 557.

[8] World Health Organization (WHO) et al. “Indoor Air Pollutants Exposure and Health Effects Report on a WHO Meeting Nördlingen, 8–11 June 1982”. In: *EURO reports and studies* 78 (1982).

[9] Pamela H Dalton and Cristina Jaén. “Responses to odors in occupational environments”. In: *Current opinion in allergy and clinical immunology* 10.2 (2010), pp. 127–132.

[10] Sumedha M Joshi. “The sick building syndrome”. In: *Indian journal of occupational and environmental medicine* 12.2 (2008), p. 61.

[11] J Madureira et al. “Indoor air quality in schools and health symptoms among Portuguese teachers”. In: *Human and Ecological Risk Assessment* 15.1 (2009), pp. 159–169.

[12] Joana Madureira et al. “Indoor air quality in schools and its relationship with children’s respiratory symptoms”. In: *Atmospheric Environment* 118 (2015), pp. 145–156.

[13] C Kielb et al. “Building-related health symptoms and classroom indoor air quality: a survey of school teachers in N ew Y ork S tate”. In: *Indoor air* 25.4 (2015), pp. 371–380.

[14] Patrizia Urso et al. “Identification of particulate matter determinants in residential homes”. In: *Building and Environment* 86 (2015), pp. 61–69.

[15] Alexander Y Mendell, Alireza Mahdavi, and Jeffrey A Siegel. “Particulate matter concentrations in social housing”. In: *Sustainable Cities and Society* 76 (2022), p. 103503.

[16] Pallavi Pant and Roy M Harrison. “Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review”. In: *Atmospheric environment* 77 (2013), pp. 78–97.

[17] Jianhua Song et al. “Prediction of pedestrian exposure to traffic particulate matters (PMs) at urban signalized intersection”. In: *Sustainable Cities and Society* 60 (2020), p. 102153.

[18] Crystal D McClure and Daniel A Jaffe. “US particulate matter air quality improves except in wildfire-prone areas”. In: *Proceedings of the National Academy of Sciences* 115.31 (2018), pp. 7901–7906.

[19] ASHRAE Standard ASHRAE. “Standard 62-2010, Ventilation for Acceptable Indoor Air Quality”. In: *American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta* (2010).

[20] Joan M Daisey, William J Angell, and Michael G Apte. “Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information”. In: *Indoor air* 13.LBNL-48287 (2003).

[21] Xiaojing Zhang, Pawel Wargocki, and Zhiwei Lian. “Human responses to carbon dioxide, a follow-up study at recommended exposure limits in non-industrial environments”. In: *Building and Environment* 100 (2016), pp. 162–171.

[22] Jan Sundell et al. “Ventilation rates and health: multidisciplinary review of the scientific literature”. In: *Indoor air* 21.3 (2011), pp. 191–204.

[23] Wen-Tien Tsai. “An overview of health hazards of volatile organic compounds regulated as indoor air pollutants”. In: *Reviews on environmental health* 34.1 (2019), pp. 81–89.

[24] Thomas Petry et al. “Human health risk evaluation of selected VOC, SVOC and particulate emissions from scented candles”. In: *Regulatory Toxicology and Pharmacology* 69.1 (2014), pp. 55–70.

[25] Maya I Mitova et al. “Human chemical signature: Investigation on the influence of human presence and selected activities on concentrations of airborne constituents”. In: *Environmental Pollution* 257 (2020), p. 113518.

[26] Tomoko Takigawa et al. “A longitudinal study of environmental risk factors for subjective symptoms associated with sick building syndrome in new dwellings”. In: *Science of the total environment* 407.19 (2009), pp. 5223–5228.

[27] J Bartzis et al. “On organic emissions testing from indoor consumer products’ use”. In: *Journal of hazardous materials* 285 (2015), pp. 37–45.

[28] João Cavaleiro Rufo et al. “Indoor air quality and atopic sensitization in primary schools: A follow-up study”. In: *Porto Biomedical Journal* 1.4 (2016), pp. 142–146.

[29] Joanna Ferdyn-Grygierek. “Monitoring of indoor air parameters in large museum exhibition halls with and without air-conditioning systems”. In: *Building and Environment* 107 (2016), pp. 113–126.

[30] Antonis A Zorpas and Antreas Skouroupatis. “Indoor air quality evaluation of two museums in a subtropical climate conditions”. In: *Sustainable Cities and Society* 20 (2016), pp. 52–60.

[31] Jonathan Williams et al. “Cinema audiences reproducibly vary the chemical composition of air during films, by broadcasting scene specific emissions on breath”. In: *Scientific reports* 6 (2016), p. 25464.

[32] Violeta Kaunelienė et al. “Indoor air quality in low energy residential buildings in Lithuania”. In: *Building and environment* 108 (2016), pp. 63–72.

[33] ZT Ai et al. “Ventilation of air-conditioned residential buildings: A case study in Hong Kong”. In: *Energy and Buildings* 127 (2016), pp. 116–127.

[34] Oliver Kinnane, Derek Sinnott, and William JN Turner. “Evaluation of passive ventilation provision in domestic housing retrofit”. In: *Building and Environment* 106 (2016), pp. 205–218.

[35] Chien-Cheng Jung et al. “Indoor air quality varies with ventilation types and working areas in hospitals”. In: *Building and Environment* 85 (2015), pp. 190–195.

[36] Tunga Salthammer et al. “Children’s well-being at schools: Impact of climatic conditions and air pollution”. In: *Environment international* 94 (2016), pp. 196–210.

[37] Mari Turunen et al. “Indoor environmental quality in school buildings, and the health and wellbeing of students”. In: *International journal of hygiene and environmental health* 217.7 (2014), pp. 733–739.

[38] Gülten Güneş, Nesibe Yalçin, and Huriye Çolaklar. “Investigation of indoor air quality in university libraries in terms of gaseous and particulate pollutants in Bartin, Turkey”. In: *Environmental monitoring and assessment* 194.3 (2022), pp. 1–15.

[39] Arefeh Hesaraki and Sture Holmberg. “Demand-controlled ventilation in new residential buildings: consequences on indoor air quality and energy savings”. In: *Indoor and Built Environment* 24.2 (2015), pp. 162–173.

[40] Ayesha Asif, Muhammad Zeeshan, and Muhammad Jahanzaib. “Indoor temperature, relative humidity and CO2 levels assessment in academic buildings with different heating, ventilation and air-conditioning systems”. In: *Building and Environment* 133 (2018), pp. 83–90.

[41] Jovan Pantelic et al. “Personal CO 2 cloud: laboratory measurements of metabolic CO 2 inhalation zone concentration and dispersion in a typical office desk setting”. In: *Journal of exposure science & environmental epidemiology* 30.2 (2020), pp. 328–337.

[42] Usha Satish et al. “Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance”. In: *Environmental health perspectives* 120.12 (2012), pp. 1671–1677.

[43] Joseph G Allen et al. “Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments”. In: *Environmental health perspectives* 124.6 (2016), pp. 805–812.

[44] Xiaojing Zhang et al. “Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance”. In: *Indoor air* 27.1 (2017), pp. 47–64.

[45] Charles J Weschler. “Roles of the human occupant in indoor chemistry”. In: *Indoor air* 26.1 (2016), pp. 6–24.

[46] Dimitrios Kotzias. “Built environment and indoor air quality: The case of volatile organic compounds”. In: *AIMS Environmental Science* 8.2 (2021), pp. 135–147.

[47] Nijing Wang et al. “Emission rates of volatile organic compounds from humans”. In: *Environmental Science & Technology* 56.8 (2022), pp. 4838–4848.

[48] Beiyu Lin et al. “Analyzing the relationship between human behavior and indoor air quality”. In: *Journal of Sensor and Actuator Networks* 6.3 (2017), p. 13.

[49] Jie Gao, Pawel Wargocki, and Yi Wang. “Ventilation system type, classroom environmental quality and pupils’ perceptions and symptoms”. In: *Building and Environment* 75 (2014), pp. 46–57.

[50] Alan Wang et al. “The Hitchiker’s Guide to Successful Living Lab Operations”. In: *arXiv preprint arXiv:2212.00008* (2022).

[51] Barbara Hoffmann et al. “WHO air quality guidelines 2021–aiming for healthier air for all: a joint statement by medical, public health, scientific societies and patient representative organisations”. In: *International journal of public health* 66 (2021), p. 1604465.

[52] David P Wyon. “The effects of indoor air quality on performance and productivity”. In: *Indoor air* 14.7 (2004), pp. 92–101.

[53] ASHRAE. “ASHRAE Handbook— Fundamentals”. In: *Atlanta, GA* (2013b).

[54] ASHRAE. “ASHRAE Standard 62.1-2013. Ventilation for Acceptable Indoor Air Quality.” In: *Atlanta, GA* (2013a).

[55] Mahsa Pahlavikhah Varnosfaderani, Arsalan Heydarian, and Farrokh Jazizadeh. “Using Statistical Models to Detect Occupancy in Buildings through Monitoring VOC, CO _2, and other Environmental Factors”. In: *arXiv preprint arXiv:2203.04750* (2022).

[56] Simona D'Oca, Tianzhen Hong, and Jared Langevin. "The human dimensions of energy use in buildings: A review". In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 731–742.

[57] Pujo Satrio et al. "Optimization of HVAC system energy consumption in a building using artificial neural network and multi-objective genetic algorithm". In: *Sustainable Energy Technologies and Assessments* 35 (2019), pp. 48–57.

[58] Chenli Wang, Kaleb Pattawi, and Hohyun Lee. "Energy saving impact of occupancy-driven thermostat for residential buildings". In: *Energy and Buildings* 211 (2020), p. 109791.

[59] Zheng Yang et al. "A systematic approach to occupancy modeling in ambient sensor-rich buildings". In: *Simulation* 90.8 (2014), pp. 960–977.

[60] Xiaohang Feng, Da Yan, and Tianzhen Hong. "Simulation of occupancy in buildings". In: *Energy and Buildings* 87 (2015), pp. 348–359.

[61] Da Yan et al. "Occupant behavior modeling for building performance simulation: Current state and future challenges". In: *Energy and buildings* 107 (2015), pp. 264–278.

[62] Farrokh Jazizadeh, Vedant Joshi, and Francine Battaglia. "Adaptive and distributed operation of HVAC systems: Energy and comfort implications of active diffusers as new adaptation capacities". In: *Building and Environment* 186 (2020), p. 107089.

[63] Wooyoung Jung and Farrokh Jazizadeh. "Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions". In: *Applied Energy* 239 (2019), pp. 1471–1508.

[64] Jonathan Brooks et al. “Energy-efficient control of under-actuated HVAC zones in commercial buildings”. In: *Energy and Buildings* 93 (2015), pp. 160–168.

[65] Avgoustinos Filippoupolitis, William Oliff, and George Loukas. “Occupancy detection for building emergency management using BLE beacons”. In: *Computer and Information Sciences: 31st International Symposium, ISCIS 2016, Kraków, Poland, October 27–28, 2016, Proceedings 31*. Springer. 2016, pp. 233–240.

[66] Zhenghua Chen, Chaoyang Jiang, and Lihua Xie. “Building occupancy estimation and detection: A review”. In: *Energy and Buildings* 169 (2018), pp. 260–270.

[67] Han Zou et al. “WinLight: A WiFi-based occupancy-driven lighting control system for smart building”. In: *Energy and Buildings* 158 (2018), pp. 924–938.

[68] Luis M Candanedo and Véronique Feldheim. “Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models”. In: *Energy and Buildings* 112 (2016), pp. 28–39.

[69] M Varnosfaderani and M Aghajani-Delavar. “Enhancing Building Information Modeling by using IoT services”. In: *Proceedings of theInternational Conference on Contemporary Iran on Civil Engineering. Academic Press*. 2017.

[70] Lars Zimmermann, Robert Weigel, and Georg Fischer. “Fusion of nonintrusive environmental sensors for occupancy detection in smart homes”. In: *IEEE Internet of Things Journal* 5.4 (2017), pp. 2343–2352.

[71] Han Zou et al. “Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT”. In: *Energy and Buildings* 174 (2018), pp. 309–322.

[72] Pawalai Kraipeerapun and Somkid Amornsamankul. “Room occupancy detection using modified stacking”. In: *Proceedings of the 9th International Conference on Machine Learning and Computing*. 2017, pp. 162–166.

[73] Dominik Scherer, Andreas Müller, and Sven Behnke. “Evaluation of pooling operations in convolutional architectures for object recognition”. In: *International conference on artificial neural networks*. Springer. 2010, pp. 92–101.

[74] Safwan Mahmood Al-Selwi et al. “RNN-LSTM: From applications to modeling techniques and beyond—Systematic review”. In: *Journal of King Saud University-Computer and Information Sciences* (2024), p. 102068.

[75] Giorgio Mustafaraj, Jie Chen, and Gordon Lowry. “Development of room temperature and relative humidity linear parametric models for an open office using BMS data”. In: *Energy and Buildings* 42.3 (2010), pp. 348–356.

[76] H Uchida Frausto, JG Pieters, and JM Deltour. “Modelling greenhouse temperature by means of auto regressive models”. In: *Biosystems Engineering* 84.2 (2003), pp. 147–157.

[77] Giorgio Mustafaraj, Gordon Lowry, and Jie Chen. “Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office”. In: *Energy and Buildings* 43.6 (2011), pp. 1452–1460.

[78] Hui Liu et al. “Intelligent modeling strategies for forecasting air quality time series: A review”. In: *Applied Soft Computing* 102 (2021), p. 106957.

[79] Ujjwal Kumar and VK Jain. “ARIMA forecasting of ambient air pollutants (O 3, NO, NO 2 and CO)”. In: *Stochastic Environmental Research and Risk Assessment* 24 (2010), pp. 751–760.

[80] Parminder Kaur, Rajwant Kaur, and Vimal Arora. “Potential indoor volatile organic compounds: Detrimental impact on health and environment”. In: *AIP Conference Proceedings*. Vol. 2558. 1. AIP Publishing LLC. 2023, p. 020009.

[81] Zaiema Rouf et al. “Volatile organic compounds emission from building sector and its adverse effects on human health”. In: *Ecological and Health Effects of Building Materials* (2022), pp. 67–86.

[82] Krati Rastogi and Divya Lohani. “An Internet of Things framework to forecast indoor air quality using machine learning”. In: *Machine Learning and Metaheuristics Algorithms, and Applications: First Symposium, SoMMA 2019, Trivandrum, India, December 18–21, 2019, Revised Selected Papers 1*. Springer. 2020, pp. 90–104.

[83] Lirong Yao and Yazhuo Guan. “An improved LSTM structure for natural language processing”. In: *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. IEEE. 2018, pp. 565–569.

[84] Jinyu Li et al. “LSTM time and frequency recurrence for automatic speech recognition”. In: *2015 IEEE workshop on automatic speech recognition and understanding (ASRU)*. IEEE. 2015, pp. 187–191.

[85] Minhao Liu et al. “SCINet: time series modeling and forecasting with sample convolution and interaction”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 5816–5828.

[86] Renzhuo Wan et al. “Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting”. In: *Electronics* 8.8 (2019), p. 876.

[87] Dongfei Yu et al. “Multi-level attention networks for visual question answering”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4709–4717.

[88] Mahsa Pahlavikhah Md Fazlay Rabbi Masum Billah. “Captioning videos using the MSR-VTT dataset”. In: (2021).

[89] Hakpyeong Kim, Taehoon Hong, and Jimin Kim. “Automatic ventilation control algorithm considering the indoor environmental quality factors and occupant ventilation behavior using a logistic regression model”. In: *Building and Environment* 153 (2019), pp. 46–59.

[90] Carrie A Redlich, Judy Sparer, and Mark R Cullen. “Sick-building syndrome”. In: *The Lancet* 349.9057 (1997), pp. 1013–1016.

[91] Gwelen Paliaga et al. “Re-Envisioning RCx: Achieving Max Potential HVAC Controls Retrofits through Modernized BAS Hardware and Software”. In: (2022).

[92] Kun Zhang et al. *Development and Verification of Control Sequences for Single-Zone Variable Air Volume System Based on ASHRAEGuideline 36*. Tech. rep. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2020.

[93] Reece Kiriu and Jeff Stein. “Medical Office Building Thrives With Advanced Control Sequences.” In: *ASHRAE Journal* 63.4 (2021), pp. 62–67.

[94] Łukasz Amanowicz, Katarzyna Ratajczak, and Edyta Dudkiewicz. “Recent Advancements in Ventilation Systems Used to Decrease Energy Consumption in Buildings—Literature Review”. In: *Energies* 16.4 (2023), p. 1853.

[95] Andrew Persily. “The Role of Carbon Dioxide in ventilation and IAQ Evaluation: 40 years of AIVC”. In: (2019).

[96] Xing Lu et al. “Advances in research and applications of CO2-based demand-controlled ventilation in commercial buildings: A critical review of control strategies and performance evaluation”. In: *Building and Environment* (2022), p. 109455.

[97] Zhe Wang and Tianzhen Hong. “Reinforcement learning for building controls: The opportunities and challenges”. In: *Applied Energy* 269 (2020), p. 115036.

[98] Davide Calı̀ et al. “CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings”. In: *Building and Environment* 86 (2015), pp. 39–49.

[99] M Humphreys. “Understanding the adaptive approach to thermal comfort, field studies of thermal comfort and adaptation”. In: *ASHRAE Technical Data Bulletin* 14.1 (1998), pp. 1–14.

[100] Rune Vinther Andersen, Bjarne Olesen, and Jørn Toftum. “Simulation of the effects of occupant behaviour on indoor climate and energy consumption”. In: *Proceedings of Clima*. Vol. 2007. 2007, 9th.

[101] Jørn Toftum. “Central automatic control or distributed occupant control for better indoor environment quality in the future”. In: *Building and environment* 45.1 (2010), pp. 23–28.

[102] Claude-Alain Roulet et al. “Perceived health and comfort in relation to energy use and building characteristics”. In: *Building research and information* 34.5 (2006), pp. 467–474.

[103] Hom B Rijal et al. “Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings”. In: *Energy and buildings* 39.7 (2007), pp. 823–836.

[104] Sebastian Herkel, Ulla Knapp, and Jens Pfafferott. “Towards a model of user behaviour regarding the manual control of windows in office buildings”. In: *Building and environment* 43.4 (2008), pp. 588–600.

[105] Rune Andersen et al. “Window opening behaviour modelled from measurements in Danish dwellings”. In: *Building and environment* 69 (2013), pp. 101–113.

[106] Davide Calı̀ et al. “Analysis of occupants’ behavior related to the use of windows in German households”. In: *Building and Environment* 103 (2016), pp. 54–69.

[107] In-Keun Shim et al. “Prevalence of sick building syndrome symptoms and subjective–objective indoor air quality of stores in underground shopping districts of Korea”. In: *Building and Environment* 228 (2023), p. 109882.

[108] Mahsa Pahlavikhah Varnosfaderani, Arsalan Heydarian, and Farrokh Jazizadeh. “A longitudinal study of IAQ metrics and the efficacy of default HVAC ventilation”. In: *Building and Environment* (2024), p. 111353.

[109] Sverre B Holøs et al. “VOC emission rates in newly built and renovated buildings, and the influence of ventilation–a review and meta-analysis”. In: *International Journal of Ventilation* 18.3 (2019), pp. 153–166.

[110] Dustin G Poppendieck et al. “Long term air quality monitoring in a net-zero energy residence designed with low emitting interior products”. In: *Building and Environment* 94 (2015), pp. 33–42.

[111] Miyuki Noguchi et al. “Measurements of volatile organic compounds in a newly built daycare center”. In: *International journal of environmental research and public health* 13.7 (2016), p. 736.

[112] Wanyu R Chan et al. “Analysis of US residential air leakage database”. In: (2003).

[113] Ed ASHRAE. “ASHRAE Handbook Fundamentals. Atlanta, GA, American Society of Heating”. In: *Refrigeration and Air-Conditioning Engineers* (2013).

[114] Andrew Persily. “Challenges in developing ventilation and indoor air quality standards: The story of ASHRAE Standard 62”. In: *Building and Environment* 91 (2015), pp. 61–69.

[115] Donald E Riesenberg and Joan Arehart-Treichel. “Sick building syndrome plagues workers, dwellers”. In: *JAMA* 255.22 (1986), pp. 3063–3063.

[116] William J Fisk and Arthur H Rosenfeld. “Estimates of improved productivity and health from better indoor environments”. In: *Indoor air* 7.3 (1997), pp. 158–172.

[117] Hakim-Moulay Dehbi et al. “Air pollution and cardiovascular mortality with over 25 years follow-up: A combined analysis of two British cohorts”. In: *Environment international* 99 (2017), pp. 275–281.

[118] Scott A Weichenthal et al. “Fine particulate matter and emergency room visits for respiratory illness. Effect modification by oxidative potential”. In: *American Journal of Respiratory and Critical Care Medicine* 194.5 (2016), pp. 577–586.

[119] Dorina Gabriela Karottki et al. “Indoor and outdoor exposure to ultrafine, fine and microbiologically derived particulate matter related to cardiovascular and respiratory effects in a panel of elderly urban citizens”. In: *International journal of environmental research and public health* 12.2 (2015), pp. 1667–1686.

[120] Zhisheng Li, Qingmei Wen, and Ruilin Zhang. “Sources, health effects and control strategies of indoor fine particulate matter (PM2. 5): A review”. In: *Science of the Total Environment* 586 (2017), pp. 610–622.

[121] Rui Chi et al. “Different health effects of indoor-and outdoor-originated PM2. 5 on cardiopulmonary function in COPD patients and healthy elderly adults”. In: *Indoor air* 29.2 (2019), pp. 192–201.

[122] Yang Geng et al. “The impact of thermal environment on occupant IEQ perception and productivity”. In: *Building and Environment* 121 (2017), pp. 158–167.

[123] Takashi Akimoto et al. “Thermal comfort and productivity-Evaluation of workplace environment in a task conditioned office”. In: *Building and environment* 45.1 (2010), pp. 45–50.

[124] Li Lan, Zhiwei Lian, and Li Pan. “The effects of air temperature on office workers’ well-being, workload and productivity-evaluated with subjective ratings”. In: *Applied ergonomics* 42.1 (2010), pp. 29–36.

[125] Louis Cony Renaud Salis et al. “Towards the definition of indicators for assessment of indoor air quality and energy performance in low-energy residential buildings”. In: *Energy and buildings* 152 (2017), pp. 492–502.

[126] Kevin Teichman et al. “Characterizing Indoor Air Quality Performance Using a Graphical Approach”. In: *NIST: Gaithersburg, MD, USA* (2016).

[127] Michel Cohas. *Ventilation et qualité de l'air dans l'habitat*. Editions parisiennes, 1996.

[128] AL Gadeau. "Assessment of ventilation strategies using an air quality index introduced in CLIM 2000 software". In: *Proceedings of Healthy Buildings, Espoo, Finland* 4 (1996), pp. 23–24.

[129] Sophie Castanet. "Contribution à l'étude de la ventilation et de la qualité de l'air intérieur des locaux". PhD thesis. Lyon, INSA, 1998.

[130] QUAD-BB. *Choix de paramètres de suivi de la qualité de l'air intérieur. Report T1. 2 for the project QUAD-BBC*. 2012.

[131] Jose Guillermo Cedeño Laurent et al. "Associations between acute exposures to PM2. 5 and carbon dioxide indoors and cognitive function in office workers: a multicountry longitudinal prospective observational study". In: *Environmental Research Letters* 16.9 (2021), p. 094047.

[132] Birgitta Berglund, T Lindvall, and Dietrich H Schwela. "World health organization occupational and environmental health team". In: *Guidelines for community noise* (1999).

[133] Theofanis Psomas et al. "Indoor humidity of dwellings and association with building characteristics, behaviors and health in a northern climate". In: *Building and Environment* 198 (2021), p. 107885.

[134] Chao Liu et al. "Influence of indoor air temperature and relative humidity on learning performance of undergraduates". In: *Case Studies in Thermal Engineering* 28 (2021), p. 101458.

[135] Emily R Jones et al. "Indoor humidity levels and associations with reported symptoms in office buildings". In: *Indoor air* 32.1 (2022), e12961.

[136] Kenichi Azuma et al. "A longitudinal study on the effects of hygro-thermal conditions and indoor air pollutants on building-related symptoms in office buildings". In: *Indoor air* 32.11 (2022), e13164.

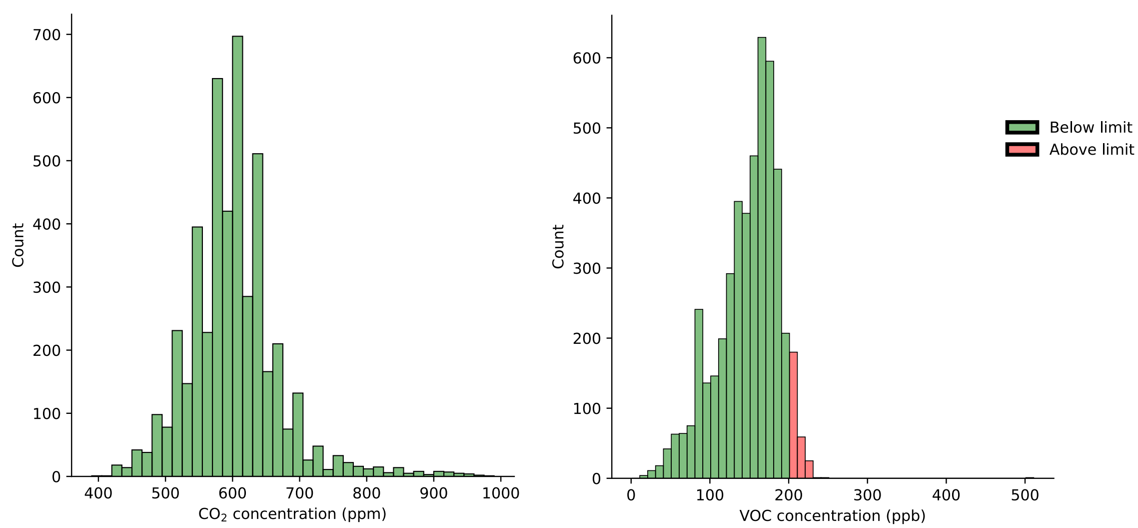
[137] Javad Razjouyan et al. "Wellbuilt for wellbeing: Controlling relative humidity in the workplace matters for our health". In: *Indoor air* 30.1 (2020), pp. 167–179.

[138] Che-Ming Chiang and Chi-Ming Lai. "A study on the comprehensive indicator of indoor environment assessment for occupants' health in Taiwan". In: *Building and Environment* 37.4 (2002), pp. 387–392.

Appendices

Appendix A

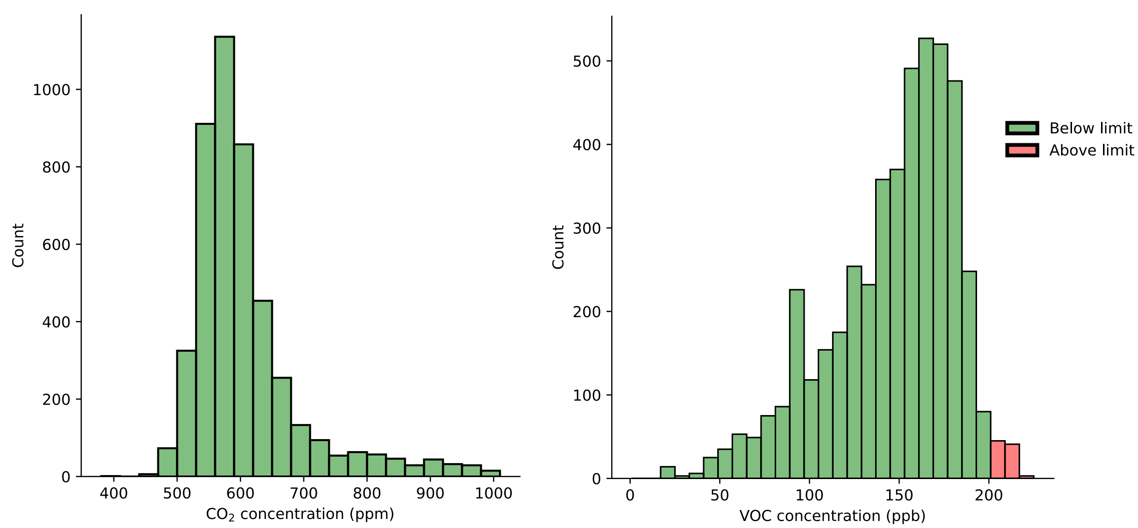
VOC and CO_2 histograms of single-occupancy of offices A and B



(a) CO_2 concentration

(b) VOC concentration

Figure A.1: Histograms of CO_2 and TVOC concentration in office A



(a) CO_2 concentration (b) VOC concentration

Figure A.2: Histograms of CO_2 and TVOC concentration in office B