EVALUATING THE IMPACT OF ENVIRONMENTAL CONDITIONS AND OCCUPANCY ON CO2 INDOOR CONCENTRATIONS THROUGH PHYSICS-BASED MODELS

By

Emma Benedek

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Civil Engineering – Master of Science

2025

ABSTRACT

Buildings account for a substantial portion of energy consumption due to heating, ventilation, and air conditioning (HVAC) contributing to a large portion of energy-related greenhouse gas emissions. Occupant behavior influences this energy usage, particularly HVAC system energy use to meet occupant needs. Thus, knowledge of occupancy and occupancy schedules can assist in more efficient building operations. While many studies examine the link between occupancy, environmental health, and energy efficiency using physics-based modeling, there have been limitations to capturing the relationship between environmental conditions and occupancy trends, and to assess their impact on model predictions and accuracy. Thus, this thesis aims to define the relationship between CO2 concentration and occupancy patterns while capturing the effect of HVAC operations and environmental conditions, in particular air exchange rates associated with changing door conditions, as well as occupancy scenarios.

To accomplish this, field data was collected on occupancy, CO2 concentrations, tracer gas testing, door state, and HVAC operations. Data was then modeled using two physics-based models for different occupant and ventilation scenarios. These models were then evaluated for accuracy under varying model assumptions, to assess the relationship between CO2 concentrations and occupancy. The evaluation included the correlation coefficient between the measured and modeled CO2 concentrations. Also, regression models were used to determine the relationship between occupancy and CO2 concentration, assessing fit using the coefficient of determination (r-squared).

Findings reflect that physics-based models can accurately determine CO2 concentrations within rooms regardless of environmental conditions and occupant trends. This further validates that physics-based models can be utilized to accurately determine CO2 concentration from occupant sources. However, findings also imply that the application of box modeling to determine occupancy trends for energy efficiency purposes based on CO2 concentration is only applicable during select conditions, limited to high rates of transient occupancy, air exchange, and unknown sources of CO2 from surrounding classrooms and hallways. This indicates that physics-based modeling is a useful tool in modeling concentrations of CO2 within spaces however should be further investigated with other aspects of VAV systems, occupancy conditions, and surrounding sources of CO2 to assist in the outcome of understanding the applicability of this model for energy efficiency purposes.

ACKNOWLEDGEMENTS

Thank you to Dr. Kristen Cetin for her guidance and support throughout my entire academic career. Her mentorship has been invaluable and instrumental in preparing me for future academic and professional endeavors. Her passion for her work has inspired and motivated me to strive for my highest potential.

I am incredibly thankful to Dr. Kaisen Lin for his guidance and support throughout my research. I greatly appreciate his dedication to air quality and pollution research, as well as his enthusiasm for teaching the subject, which had sparked my own interest.

Additionally, I would also like to thank Dr. Annick Anctil and Dr. Kaisen Lin for serving as a part of my thesis committee and providing insightful feedback and assistance. I am incredibly grateful to have such a knowledgeable group of mentors to work with and appreciate everything that you have all helped me with.

A heartfelt thank you to all of my amazing friends and family, especially Katie Toneilli, whose support in academics, work, and friendship has been invaluable to my academic journey. Lastly, a huge thank you to my dog Henry, who has been a constant companion at my desk throughout writing my entire thesis.

TABLE OF CONTENTS

INTRODUCTION

Background

Buildings account for a substantial portion of energy consumption; in 2023 commercial buildings alone accounted for 1.72×10^{19} joules ($16{,}343 \times 10^{12}$ BTUs) of energy (U.S. EIA, 2024). This energy usage contributes to a significant amount of greenhouse gas emissions, as approximately 40% of all energy-related emissions come from buildings (ASHRAE, 2022a). A large portion of these emissions are due to heating, ventilation and air conditioning (HVAC) which accounts for 36% of all commercial building energy usage (Kong et al., 2022). Occupant behavior drives much of building energy usage, particularly HVAC system energy use, which provide ventilation and temperature controls to meet occupant needs, with the goal of providing occupant comfort and a healthy indoor environment (Mannan et al., 2021; Kong et al., 2022; Bian et al., 2024). For this reason, knowledge on occupancy and occupancy schedules can be used to support more efficient operations of building system, including controls for lighting, thermostat setpoints and setbacks, outdoor air ventilation requirements, as well as the use and/or automation of operable windows and shading controls (Yan et al., 2015).

The link between occupancy, environmental health, and energy efficiency has been increasingly studied in the context of occupant-based controls of buildings (Chu et al., 2022; Kong et al. 2022; Mitra et al. 2022). Specifically, research has focuses on both developing methods of occupancy detection and/or counting, as well as research on how occupancy related information can inform controls to improve building and HVAC energy efficiency and indoor environmental quality. For occupancy detection and/or counting methods that have been developed in recent years, these have been characterized and summarized by several recent review papers (Chu et al., 2022; Li et. al. 2024; Mitra et al. 2022). Common methods discussed include the use of motion-based, radiofrequency-based, sound wave-based, camera-based, and/or infrared sensors, as well as through the use of multiple sensing methods, termed “sensor fusion”, combined with predictive models. Such models are primarily data-driven models, using statistical, regression, machine learning, and/or artificial intelligence (AI) methods. However, such data-driven methods for occupancy detection have also struggled with sensor and/or algorithm reliability challenges. This has been recognized as of sufficient importance for U.S. funding agencies to require independent testing of occupancy sensors systems that use data-driven methods as a part of occupancy sensor development (ARPA-E, 2017), and for the support of development of standard methods of testing occupancy sensor systems' performance (Kula et al., 2023; Chu et al., 2022) due in part to inconsistencies in accuracy reporting.

For example, Kong et al. (2022) demonstrated the potential of occupancy-based detection for improving HVAC energy efficiency and indoor air quality using a combination of a radiofrequency-based and two motion-based sensors. However, during the study some sensors required consistent calibration and occupancy detection reliability was low in parts of the study. Other common methods of occupancy-based detection using motion-based sensors have also been noted to have the potential for reliability issues as they do not detect periods of fine motion (Yang et al., 2018). Radiofrequency-based sensors are impacted by noise and require the occupant to have a device to transmit signals; sound wave sensors could be interfered with by clothing and materials in the facility, and some infrared sensors require constant motion with unobstructed sight and/or constant maintenance (Chu et al., 2022; Mitra et al., 2022). There are also concerns about capturing detailed and personally identifiable information through the use of camera-based methods for occupancy detection, thereby posing a heightened risk to data security and occupant privacy (Cali et al. 2014; Huang et al., 2024).

Although there are advantages to data-driven occupancy detection systems, there are also challenges. In particular there are concerns about trained algorithms being highly dependent on the environment in which they are placed, making them less useable if removed and placed in a different space (Mitra et al., 2022). This is consistent with the challenges of the use of data-driven methods that require a significant period of training data (Gu et al., 2021). In addition, if conditions change, the algorithm needs to be retrained. The ability of a data-driven model to predict conditions outside of the scenarios in which it was trained to consider is also typically limited (Miao et al., 2023). Also, if conditions change which are not captured by the independent variables included in the data-driven method(s), this can also cause discrepancies between predicted and measured values (Gu et al., 2021).

As alternative to data-driven models (black-box), the other primary method for predictive modeling is physics-based (white box) models. In the context of occupancy detection and/or counting, the primary type of physics-based model that is applicable is the use of mass balance equations of the building space, primarily associated with indoor air pollutants produced by occupants. In this case one of the primary pollutant produced reliably by occupancy through human breathing is carbon dioxide (CO2). Prior studies have used indoor CO2 concentrations as one way to detect and/or count occupancy, especially in the context of air quality and ventilation controls (Dedesko et al., 2015; Sun et al., 2011). A summary of such studies is shown in Table 1. This includes some scenarios in classroom settings. As an example, Zuraimi et al. (2017) compared the performance of physical and statistical models to predict occupant counts in a high-volume lecture hall using CO2 sensors. This study found that while the utilized physics-based models' required a larger number of inputs, these models were found to be able to adequately predict occupancy counts in a particular location when environmental factors such as air exchange rates (ACH) are accurately measured and accounted for.

In a significant number of studies, particularly those focused on indoor air quality (IAQ) and/or IEQ, mass balance modeling was used with occupant generated CO2 (Asif & Zeeshan, 2020; Chang, et al., 2009; Fan et al., 2022; Lawrence & Braun, 2007; Li et al., 2014). Other studies have utilized physics-based modeling methods for comparing the performance of multiple model types to predict occupant counts and energy performance (Zuraimi et al., 2017) and to compare measured CO2 concentrations, volume of room per person, and occupancy for optimal HVAC periods (Franco & Leccese, 2020). As summarized in this table, prior studies have evaluated this relationship in buildings with a range of HVAC system types and space sizes. Most studies also incorporate the use of tracer gas approaches to inform mass balances of CO2, as tracer gas is a commonly used method to evaluate air exchange rates (Chang et al., 2009; Fan et al., 2022, Lawrence & Braun, 2007; Zuraimi et al., 2017; Li et al., 2014).

Table 1: Literature summary of physics-based modeling of occupant-based CO2 concentrations including location, time, characteristics and the associated study

Setting	**Location**	**Room type**	**Time Frame**	**Building Characteristics**					**Performance Metric**	**Research Study**
				HVAC System	Size			Occu-pancy		
					Height (m)	Area (m^2)	Volume (m^3)			
University	Pakistan	Primary School Classrooms	4 months	Naturally ventilated with split type AC units and portable fan heaters	-	20.3 to 33.9	-	18 to 29	Pearson's Correlation Coefficient >0.98 (Minute-by-minute ventilation rate) <0.70 (Averaged ventilation rate)	Asif & Zeeshan, 2020
	Taiwan	Computer Lab and Classroom	-	Window-type and water-cooled package	-	-	296 to 595	22 to 36	One-way analysis of variance across CO2 concentra-tions measured (ANOVA) $p<0.001$	Chang, et al., 2009
	China	Research room	4 weeks	Naturally ventilated	2.8	23	64.4	0 to 7	Maximum nonuniformity coefficient 5.68%	Fan et al., 2022

Table 1 (cont'd)

	Italy	University classrooms	4 months	Natural ventilation	-	73 to 336	212 to 1587	72 to 366	Linear trend of CO2 compared to volume of space for each occupant	Franco & Leccese, 2020
	China	Dormitory	12 months	Naturally ventilated	3	17.1	51.3	4	Relative error in ACH 1.6% to 7.8%	Li et al., 2014
	Taiwan	Lecture theatre	4 months	Constant air volume system with a variable speed drive	-	-	876	Less than 200	Correlation coefficient 0.80 to 0.97	Zuraimi et al., 2017
Commercial Buildings	California	Restaurants and School Classrooms	Several months to 1 year	Demand-controlled ventilation	-	75 to 835	-	Vari-able	Coefficient of variation 4% to 15%	Lawrence & Braun, 2007

However, limitations of current research include that most studies have used buildings with natural ventilation (Fan et al., 2022; Franco et al., 2022; Li et al., 2014) and/or relatively short-term datasets, with most ranging from 4 weeks to 4 months (Asif & Zeeshan, 2020; Fan et al., 2022; Franco & Leccese, 2020). Additionally, it was noted in these studies that results can be influenced by multiple factors such as the opening and closing of doors and windows and variations in occupants. However these factors have not been significantly studied. In addition, most studies that have compared occupancy and CO2 concentrations use either scenarios with near-constant occupancy (Chang, et al., 2009; Li et al., 2014) or transient occupancy (Asif & Zeeshan, 2020; Fan et al., 2022; Franco & Leccese, 2020; Lawrence & Braun, 2007; Zuraimi et al. 2017). None have evaluated the use of physics-based models that support the prediction of the relationship between CO2 concentrations and occupancy across a range of types of increasing, decreasing, and near constant occupancy scenarios all in the same environment or using the same modeling framework.

Specifically, Asif et. al (2020) calculated the ventilation rate for classroom spaces by averaging the change of occupant-generated CO2 concentration over time. While variation in the envelope, doors/windows openings, were included in this average, the direct impact from these specific parameter's ventilation variation was not analyzed separately. In the results it was noted that the CO2 concentration calculated by averaged ventilation rate compared to measured data resulted in lower performance and the approach could be improved. Furthermore, when the tracer-gas-concentration decay method to calculate ventilation rate, Chang, et al. (2009), did not account for changing ventilation rates within the studied space or mention accounting transient occupant conditions. Additionally, in Li et al. (2014) found infiltration rates did not significantly affect CO2, however there were limited times when doors/windows were opened. In Zuraimi et al. (2017) it was noted that the model accuracy was affected by air exchange rates, however this was just noted to be caused by the method by which sensors collected data affecting how responsive the model was. Additionally, in Franco & Leccese (2020) environmental factors were also noted to be important to support mass balance modeling of occupant-based CO_2 concentrations. It was determined that CO_2 was directly related to occupancy but the initial CO_2 concentration, and the volume of the room, occupancy count, and air exchange rate were all influential in the results. Notably the openings of doors/windows were mentioned to influence results, however quantifying their impact was determined to be challenging.

In summary, as noted in Mitra et al. (2022), methods of occupancy modeling are influenced by the space within which they are created, and environmental changes can also impact performance. In addition, as stated in the ASHRAE Position Document on Indoor Carbon Dioxide, it is essential to assess factors that influence CO_2 concentrations within a space (ASHRAE, 2020b). Therefore, further analysis and testing is needed to assess this relationship including the effect of varying air exchange rates corresponding to variation in opening and closing of doors/windows and HVAC operation schedules, as well as varying occupancy conditions, and to assess the impact of the environment on physics-based model predictions and accuracy.

This study thus aims to define the relationship between CO_2 concentration and occupancy patterns and capture the effect of varying HVAC operations, and varying environmental conditions, in particular air exchange rates associated with changing door conditions, as well as varying occupancy scenarios. To accomplish this, field data was collected on occupancy, CO_2 concentrations, tracer gas test results, door state, and HVAC operations, which is then utilized as input into physics-based box models. These models are then evaluated for accuracy under varying environmental conditions and model assumptions, to assess the relationship between CO_2 concentrations and occupancy.

The remainder of this research is organized as follows, the methods section discusses the monitored space, data collection, field testing, the developments of models for analysis, and the methods for analysis. The results section details model performance and the relationship with the concentration of CO_2 predicted in models to occupancy. The final section includes conclusions and future work for the study.

METHODS

Overview

To evaluate the relationship between CO2 and occupancy under variable conditions in an occupied building and the factors that influence this relationship, field data was collected at 1-minute intervals between June and November 2022. This data included CO2 concentrations at multiple locations within the space utilized, count of occupants, and building and systems operations data. This data was then used to develop and validate a physics-based box model of the tested space.

Description of Monitored Space Used for Field Data Collection

Data was collected from an academic building classroom space located in ASHRAE Climate Zone 5A in the Midwest area of the United States. The classroom size is approximately 8.2 m (27 ft) by 9.7 m (32 ft) with a ceiling height of 3 m (10 ft), as shown in Figure 1. This classroom has a rated occupant capacity of 52 people and two interior doors leading to an adjacent hallway. The classroom is located on the exterior perimeter of the 23,000 m^2 (250,000 sq ft) building, on the first floor, with two other classrooms on either side. The ceiling is a drop ceiling with a plenum space above, which is connected to a return air duct. This space was chosen for field data collection as it was occupied regularly, typically multiple times per day, between the hours of operation of the building from 7:00 A.M. to 7:00 P.M Monday through Friday in both the fall (August to December) and spring (January to May) semesters.

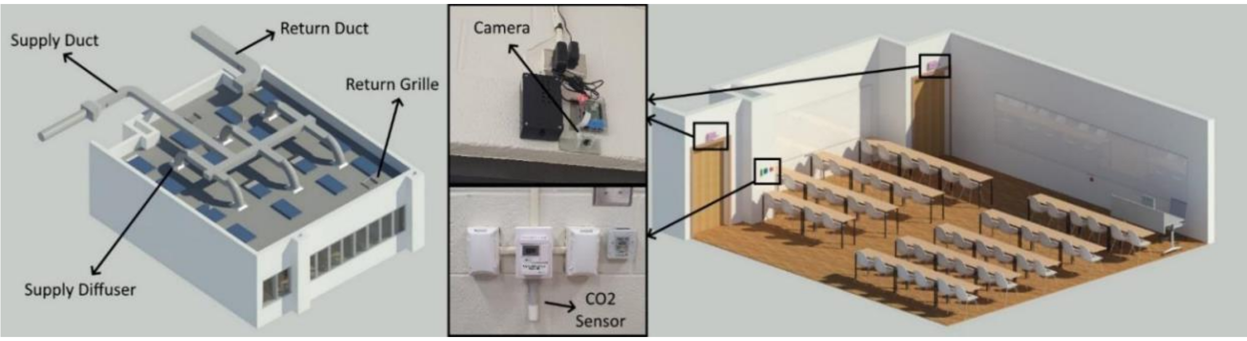

Figure 1: Exterior view (left) and interior view (right) of the field-testing location, including CO2 sensors and occupancy cameras (middle)

The heating, ventilation, and air conditioning (HVAC) servicing this room included an air handling unit (AHU) that serviced this room and three other adjacent classroom spaces on the same floor. The AHU included a mixing box that allows for some of the return air from the classroom spaces to be mixed with outdoor air, conditioned, and returned to the spaces as supply air. The remaining return air was exhausted to the exterior. Once the supply air was ducted through the supply air ducts, the amount of supply air provided to the classroom was determined via a VAV box with a damper adjusted using temperature-based controls. This VAV box controlled supply air to six supply air diffusers evenly spaced throughout the classroom ceiling. When the temperature of the room reached above the set threshold, the damper was opened automatically to increase the supply airflow of conditioned air. Once the target temperature was reached, this damper closed automatically to accommodate the minimum air flow rate. Return air sent back to the AHU was received through two return air grilles which open to a plenum space above the drop ceiling. This plenum is connected to a return air duct.

The observed setpoints used in the test space ranged from 20.5 C (69 F) to 23.3 C (74 F) throughout the test period. Temperatures were observed to be at a minimum of 20.5 C (69 F) during unoccupied hours and peaked at 23.3 C (74 F) during periods of high occupancy, with at least 30 occupants in the room. During the occupied periods the room most commonly had 1-3 or 16-18 occupants, and temperatures ranged from 21.1 C (70 F) to 22.7 C (73 F). The observed supply air volume ranged from 0 m^3/s to 0.25 m^3/s (548 cfm) during occupied periods. The air flow sensor installed in the VAV box used for these measurements was calibrated prior to the testing period.

Long Term CO2, HVAC, and Occupancy Data Collection

To monitor CO2 concentrations in this space, multiple commercially available CO2 sensors were installed in various locations throughout the room, including Vaisala GMP252 Probe (Vaisala, 2023) , Aranet4 Pro (SAF Tehnika JSC, 2025), ACI A/CO2-R2 wall sensor (Automation Components, Inc., n.d.), and in the supply and return air ducts, Vaisala GMP252 Probe (Vaisala, 2023), as shown in Figure 2. Sensors were calibrated weekly throughout the testing period using standard CO2 gas cylinders at concentrations of 10 ppm, 400 ppm, and 1000 ppm. The supply and return air duct CO2 sensors' data was collected via connection to the building management system (BMS), as were several of the room sensors; the remaining were collected wirelessly via Bluetooth every approximately 2 weeks, then combined to create a complete dataset. Figure 2 provides a diagram showing where each of these sensors were in the test space. For those located inside the classroom, Table 2 indicates their height above the floor. Table 3 provides the manufacturer-reported specifications for all sensors.

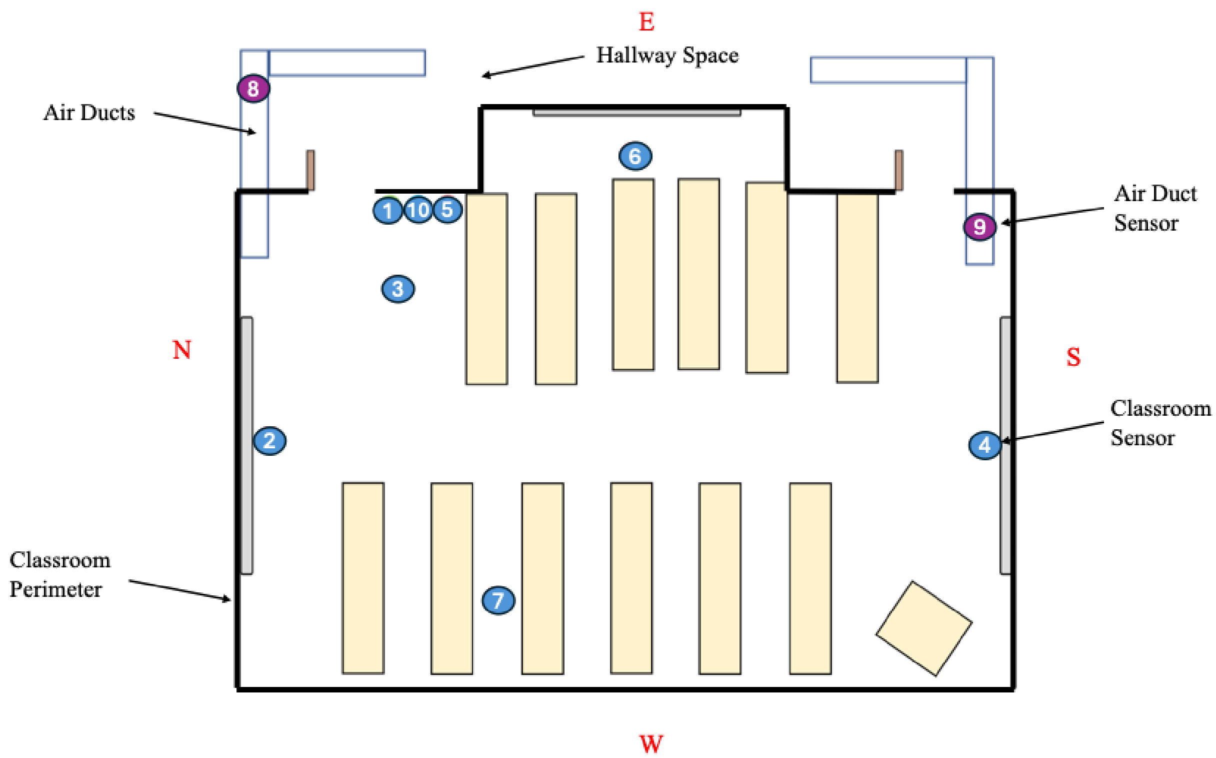

Figure 2: Top-down view of the classroom sensor placements are as follows: *Sensor 1* – East wall beside right door of classroom; *Sensor 2* – North wall of classroom; *Sensor 3* – Northeast ceiling mounted next to the supply air duct; *Sensor 4* – South wall of classroom; *Sensor 5* – East wall beside right door of classroom; *Sensor 6* – East ceiling of classroom; *Sensor 7* – West ceiling of classroom

Sensor 1, located next to one of the entry doors into the classroom from hallway, and *Sensors 8 and 9* located in the supply and return air ducts, respectively, were all the same brand of sensor, and were selected based on comprehensive comparative testing that evaluated how well the sensors performed under a range of conditions (Cetin et al. 2024). *Sensors 2-7* were placed at various locations on the walls and ceiling of the classroom to assess the spatial distribution in CO2 concentration and level of mixing throughout the space. Throughout the monitoring period, the CO2 concentrations measured from the supply air duct ranged from 423 ppm to 528 ppm, while that for return air duct ranged from 408 ppm to 1099 ppm. CO2 concentrations in the classroom ranged from 415 ppm to 1289 ppm across all the tested sensors.

Table 2: Location and height of placed CO2 sensors

Sensor Name	Location	Approximate Height of Sensor (from Ground level)
Sensor 1	East wall beside right door of classroom	1.2 m (4 ft)

Table 2 (cont'd)

Sensor 2	North wall of classroom	0.6 m (2 ft)
Sensor 3	Northeast ceiling mounted next to the supply air duct	3 m (10 ft)
Sensor 4	South wall of classroom	0.6 m (2 ft)
Sensor 5	East wall beside right door of classroom	1.2 m (4 ft)
Sensor 6	East ceiling of classroom	3 m (10 ft)
Sensor 7	West ceiling of classroom near classroom window	3 m (10 ft)

Table 3: Manufacturer-reported CO2 and temperature sensor accuracy and range

<table>
<tr><th>Sensor Name</th><th>CO2 Accuracy</th><th>CO2 Range</th><th>Temperature Accuracy</th><th>Temperature Range</th></tr>
<tr><td>Sensor 1, 8, 9</td><td>± 40 ppm</td><td>0 ppm - 5000 ppm</td><td>± 0.5°C</td><td rowspan="2">0°C – 50 °C</td></tr>
<tr><td>Sensor 2, 3, 4, 5, 6 ,7</td><td>± 30 ppm</td><td>0 ppm - 9999 ppm</td><td>± 0.3°C</td></tr>
</table>

To track occupancy and the opening and closing of the hallway doors, two camera modules connected to custom programmed microcontrollers (Raspberry Pi Foundation, 2024) and were placed above the room's doorways as shown in Figure 1, such that the cameras could only see the top of each person's head that was entering or exiting. Videos were recorded throughout the testing period for both doorways via the use of a Python code developed and used to store daily videos with timestamps from each camera. These video files were downloaded weekly then used to manually count the number of occupants in the space. A spreadsheet was generated for each door, and every time a person entered or exited through a door, it was manually recorded in the spreadsheet along with the corresponding timestamp.

After, the data for each door were compiled to compute the occupancy of the classroom with a 1-minute frequency. In addition, the video data was also used to determine during which periods the door(s) were open or closed. The opening and closing of the doors were important to document in order to evaluate how this change in state impacted air exchange rates in the test space, and thus relationship between occupancy and CO2 concentrations. Based on the data collected the occupied periods only occurred between the hours of 7:00 A.M. to 7:00 P.M on weekdays (M-F), which is associated with when the classroom was unlocked. Considering only the times the classroom was open, it was occupied approximately 33% of the time across the time of data collection. For the periods where the room was occupied, the distribution of number of occupants is provided in Figure 3. The classroom occupancy was mainly occupied by 1-3 people (34%) and 16-18 people (11%) over the monitoring period. For the periods where there were a smaller number of people, the space was generally being used for cleaning or studying; for the periods with a larger number of people, the space was being used as a classroom.

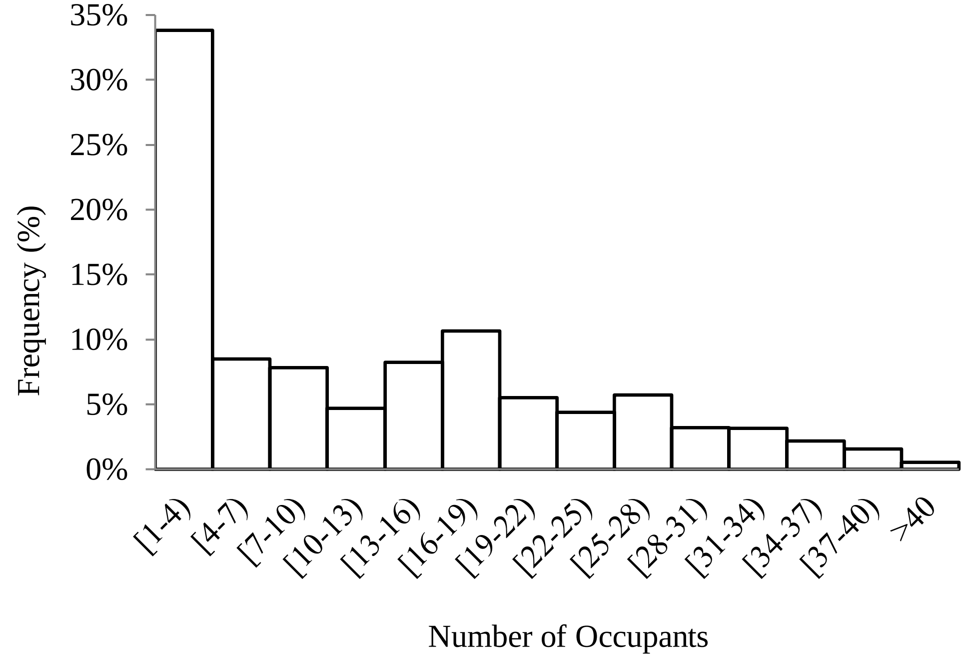

Figure 3: Occupancy distribution of classroom space excluding unoccupied times

Field Data Quality Control & Subdivision of Scenarios

Once all data was compiled into a single combined dataset, the data was quality controlled across all variables to determine which periods of data across the monitoring period were useable, free of erroneous data, and included a complete set of data across all variables. Any instances of missing data arising from factors such as sensor removal for calibration, video errors, or other scenarios, were removed from the final dataset. Unoccupied intervals shorter than 30 minutes were also excluded from the dataset. Additionally, time periods with highly variable occupancy with frequent door openings (occurring in intervals shorter than 30 minutes) were also excluded. The distribution of CO2 concentrations was also reviewed to assess for anomalies or outliers; however, none were identified. In total the final dataset includes 1,661 minutes of data, which was used for analysis.

The data was then sorted into days where the occupancy was greater than zero at some point during the day, and other days where no occupancy occurred. Days with no occupancy were also removed from the dataset. Periods of time where occupancy occurred were then further divided into multiple scenarios. Division of the data first included dividing the data into periods where the (i) 0, (ii) 1 or (iii) both classroom-hallway doors were open (3 scenarios). Door conditions were documented in the datasheets for all occupied periods. Second, the data was further subdivided to account for the influence of changes in occupancy across the test period. Three scenarios were considered: (i) steadily decreasing occupancy, (ii) steadily increasing occupancy, and (iii) near constant occupancy. The data was divided in this way to analyze how varying occupancy trends impact the accuracy of modeling CO2 in results. Ideally CO2 should be released and analyzed under steady state conditions however periods of rapid transient conditions have a potential impact on the rate at which CO2 is released.

Natural Air Exchange Rate

To evaluate the natural air exchange rate of the test space when the HVAC system was turned off, a tracer gas test method was used. This is a commonly method to evaluate air exchange, including in university settings in classrooms (Chang et al., 2009; Fan et al., 2022; Li, et al., 2014). ASTM E741-23 standard was followed to complete this testing. Initially, a blower door test was performed, however, due to the presence of the plenum space and drop ceiling, the blower door test could not reach the targeted differential pressure for the system to function while pressurizing the room. Ceiling tiles were pushed up due to the pressure difference caused by the blower door in the room. Thus, instead the tracer gas testing method was used.

To ensure accurate testing the space and concentration of the tracer gas should be uniform (ASHRAE, 2022b; ASTM International, 2023). Multiple tests under various environmental conditions were conducted in October 2023, January 2024 and February 2024. To ensure no external influence from surrounding classrooms, testing was performed during a time when no class sessions were scheduled in the room or the surrounding classrooms serviced by the same air handling unit. Additionally, to verify the room's natural air exchange rate was influencing the air exchange of CO2 in the room, the building management system (BMS) system was used to verify that the HVAC system was set to the minimum supply air flow during testing dates. This was verified from the calculation of the air exchange rate based on the average air volume entering the room from the HVAC system supply duct. HVAC air volume was determined as negligible during testing as air exchange rates were, on average, 0 ACH or 1.2-1.5 orders of magnitude lower than the calculated air exchange rate from the tracer gas test (see Appendix Table A.1).

Five pounds of dry ice was used as the source of CO2 for each test. Previous studies suggest that the uncertainty of this method in determining the air exchange rate is less than 10% (Cheng & Li, 2014). All CO2 sensors located in the room were monitored for 20-30 minutes prior to the start of testing to confirm an even distribution of CO2 and to establish baseline conditions. Then, dry ice was placed in the classroom to achieve a targeted maximum concentration of 2000 ppm. Dry ice was equally spaced in three styrofoam containers on three tables around the classrooms with fans placed next to the dry ice in an effort to evenly disperse the CO2. Once the target CO2 concentration was reached, the dry ice containers were capped to eliminate further release of CO2. The level of CO2 in the classroom was monitored and measured for up to 2 to 3 hours until CO2 concentrations returned to baseline levels that occurred before testing began.

Three different air exchange rates were measured to ensure a range of anticipated conditions were captured including (i) both doors to the hallway closed, (ii) one door open and one closed, and (iii) both doors open. For the scenario where all doors were shut, during testing all doors were shut and CO2 levels were actively monitored until the room achieved approximately 2000 ppm. The dry ice and fans were then removed from the room and the concentration of CO2 while it decayed was collected. For the scenario where all doors where opened, during testing all doors were left open and the previous steps were repeated. This was also repeated for conditions where the left or right doors were open/shut. Data was then analyzed using the linear regression of the change in rate of CO2 over time (Equation 1). The time interval plot equation is as follows:

$$C_t = C_a * (1 - \exp(-I_{natural} * t)) + C_{t=0,max} * \exp(-I_{natural} * t) \qquad (1)$$

Where C_t is the concentration of CO2 at time t. C_a is the concentration of the pollutant can be represented as the ambient concentration in the room $\left(\frac{mg}{m^3}\right)$. $I_{natural}$ is the infiltration rate from just the building envelope in air changes per hour (ach). $C_{t=0,max}$ is the maximum concentration of CO2 once the dry ice was removed from the room. Time, t is the time interval since the occurrence of $C_{t=0,max}$.

Data was first graphed and regressed separately for *Sensors 2-7* to determine if a sensor was an outlier from the system and did not follow the other sensor trends or if there was incomplete mixing in the room. The air exchange rate was determined based on the slope of the linear regression model, reflecting the change in concentration of CO2 overtime. Data was then averaged from the sensors with similar calculated air exchange rates. For each field test, all data used to determine air exchange results was used as input into a linear regression except for *Sensor 3* for all scenarios and *Sensor 6* (for one door open and one closed). These were not used because their values had a significant difference compared to all other sensors used to calculate the air exchange rates for the same time interval (Appendix Table A.1). Sensor *3* was located closest to the supply air duct, thus was likely overly influenced by the minimum air flow CO2 concentrations from the supply air duct; *Sensor 6* was directly above one of the containers of dry ice, resulting in higher CO2 concentrations than other sensors during some scenarios. The average of the calculated air exchange rate was used for the four tested scenarios, doors open, left door open, right door open, and doors closed.

Physics-based Box Model Development

To develop a physics-based model of the test space, for use with the field collected data for validation, a box model was used. This model was created under the assumption that the room was a well-mixed space, where CO2 was uniformly distributed from the source. This was verified through the comparison of *Sensors 2-7* CO2 measurements which demonstrated that all CO2 concentrations were within 33% when HVAC is on and 32% when the HVAC system was off. It should be noted that the HVAC system was off (no flow) during the analysis period for approximately 5% of the time and on, at varying levels of generally low air flow rates, for 95%. It was assumed that CO2 generated from occupants would be a uniform source distributed throughout the classroom. This model was based upon the mass-balance principle, the rate of increase/decrease in concentration of CO2 in the room was equivalent to the difference between the mass rate of CO2 being generated from students and exiting, under the assumption that CO2 will not decay, expressed as Equation 2 (Asif & Zeeshan, 2020).

$$\frac{VdC}{dt} = Q_{in}(\mathrm{C}_{in}) - Q_{out}(\mathrm{C}_{out}) + E \quad (2)$$

Whereas V represents the volume of the room (m^3). $\frac{dC}{dt}$ is the change of pollutant concentration in respect to time. C_{in} is the concentration of pollutant entering the room from both natural and mechanical ventilation ($\frac{mg}{m^3}$). Q_{in} is the flow rate of pollutant into/out of the room $\left(\frac{m^3}{s}\right)$. E is the source of pollutant (number of persons (n) * generation rate ($\frac{mg}{s} * person$). The generation rate of CO2, n, is based upon the number of people in the room from the collected occupancy data. The number of people is then multiplied by the average generation rate of CO2 per person, $4.25 * 10^{-6} \frac{m^3}{s*person}$ ($8.3 \frac{mg}{s*person}$) . This was determined based on the assumption that the classroom demographic was 50% male and 50% female from 20-29, mainly sitting and/or typing, therefore the average of the two CO2 generation rates under each category were taken (Yang et al., 2020). C_{out} is the concentration of CO2 in the room $\left(\frac{mg}{m^3}\right)$. Whereas, $Q = I * V$ (given I is the infiltration rate in air changes per hour (ach) and V is the volume per time ($\frac{m^3}{s}$) and $C_{out} = C_a$(where $C_a$is the concentration of the pollutant can be represented as the ambient concentration in the room $\left(\frac{mg}{m^3}\right)$. When integrated to express concentration as a function of time (from time *0* to time *t*) this equation was rewritten as Equation 3:

$$C(t) = \frac{E/_V + C_a I_x}{I_x} (1 - e^{-I_x t}) + C_{(t-1)} e^{-I_x t} \quad (3)$$

1-minute frequency was used for all variables to increase accuracy as CO2 concentrations, amount of people in the room which impacts the source, and the environmental conditions were observed to change minute-by-minute. $C_{(t-1)}$ is the concentration of CO2 determined by the previous timestep. I_x is used to represent the mechanical (I_{supply}) and/or natural ($I_{natural}$) air exchange rate depending on the scenario considered.

Each scenario considers different sources of ventilation within the room. Scenario 1 represents when the HVAC system's supply air into the room is significantly higher than the natural air exchange rate determined by the field test. In this case the air exchange rate is only represented by the volume of air being supplied into the room by the HVAC system. Scenario 2 represents when the natural air exchange rate is significantly higher than the mechanical air exchange rate. Here, the air exchange rate variable is only represented by the air exchange rate determined in the field test. Scenario 3 represents a mix of ventilation, when the rate of air infiltration was approximately equivalent to the mechanical ventilation rate. Each scenario was compared based on the calculated correlation coefficient between the scenario and the measured CO2 concentrations from the return duct, as discussed below.

Scenario 1, Mechanical Ventilation Dominated

The first scenario represents scenarios where the influence of the HVAC system's operation and air exchange rate is greater than that of the natural air exchange rate of the room, thus the model of the room can be represented by the HVAC system's operation only. For the first model, Equation 3 is used and the air exchange rate I_{supply} is the volume of air being supplied by the HVAC system. Figure 4 represents the box model of Scenario 1 and Scenario 2.

Scenario 2, Natural Ventilation Dominated

In the second scenario, Equation 3 is also used. In this scenario, the influence of the natural air exchange rate of the room is dominate over HVAC operations. The air exchange rate is represented by $I_{natural}$, as determined through tracer gas testing. This scenario represents when the HVAC system is off or very low and/or with relatively low CO2 concentrations in the supply air. Depending on if no, one or two doors were open this air exchange rate is adjusted throughout the monitoring period.

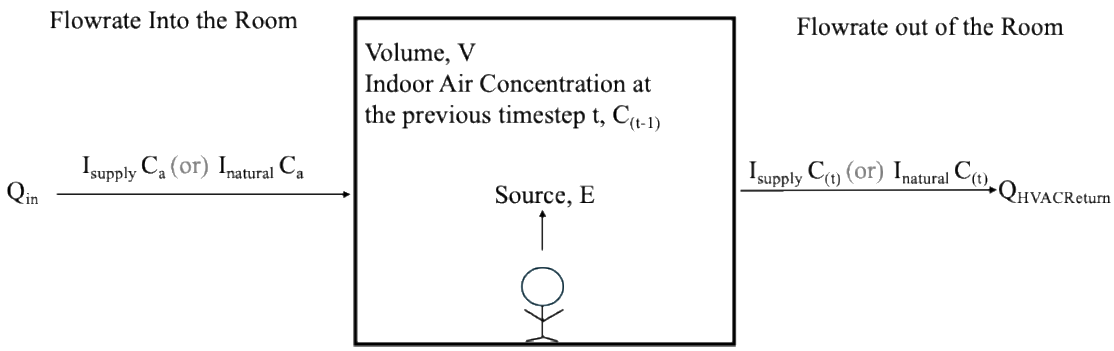

Figure 4: Box model of Scenario 1 and 2 mechanical or natural dominated air exchange

Scenario 3, Mixed Mechanical and Natural Ventilation

In the third scenario, this model represents when both the supply air flow from the HVAC system and the natural air exchange rate must both be accounted for, for example if both air exchange rates were a similar level of magnitude with similar levels of CO2 concentrations, and thus important to account for both. For this scenario a different equation is derived. The flowrate into the room can be represented using Equation 4.

$$Q_{total} = Q_{HVAC\ Return} = Q_{supply} + Q_{natural} \qquad (4)$$

Where Q_{total} is the total flowrate into the room $(\frac{m^3}{s})$. $Q_{HVAC\ Return}$ is the HVAC return flowrate out of the room $\left(\frac{m^3}{s}\right)$. Q_{supply} represent the flow rate of the supply air into the room $(\frac{m^3}{s})$. $Q_{natural}$ is the flow rate of the air due to the natural air exchange rate $\left(\frac{m^3}{s}\right)$. Given that $I = \frac{Q}{V}$, the following equation for the mass balance of the room including both mechanical and natural air exchange can be rewritten as the follows, as shown in Equation 5.

$$\frac{VdC}{dt} = (\mathrm{I}_{supply}\,\mathrm{C}_{supply} + \mathrm{I}_{natural}\,\mathrm{C}_{natural})\mathrm{V} - (\mathrm{I}_{HVAC\ Return}\mathrm{C}_{HVAC\ Return})\mathrm{V} + S \qquad (5)$$

And when integrated to express concentration as a function of time (from time 0 to time *t*) this equation could be rewritten as Equation 6:

$$C(t) = \left(\frac{E}{V\,\mathrm{I}_{total}} + \frac{(\mathrm{I}_{supply}\mathrm{C}_{supply} + \mathrm{I}_{natural}\,\mathrm{C}_{natural})}{\mathrm{I}_{total}}\right)(1 - \mathrm{e}^{-\mathrm{I}_{total}t}) + C_{(t-1)}e^{-\mathrm{I}_{total}t} \quad (6)$$

Whereas $I_{total} = I_{supply} + I_{natural}$ (given I is the infiltration rate in air changes per hour (ach) for both supply and natural values). I_{supply} represented the HVAC air flow rate data (ach) and $I_{natural}$ represents the natural air exchange rate, depending on the number of doors to the hallway that are open. C_{supply} is the concentration of CO2 in the supply air $\left(\frac{mg}{m^3}\right)$. $C_{natural}$ is assumed to be 420 ppm, the average outdoor air concentration of CO2 (ASHRAE 2022b.). See Figure 5 which represents the box model of Scenario 3.

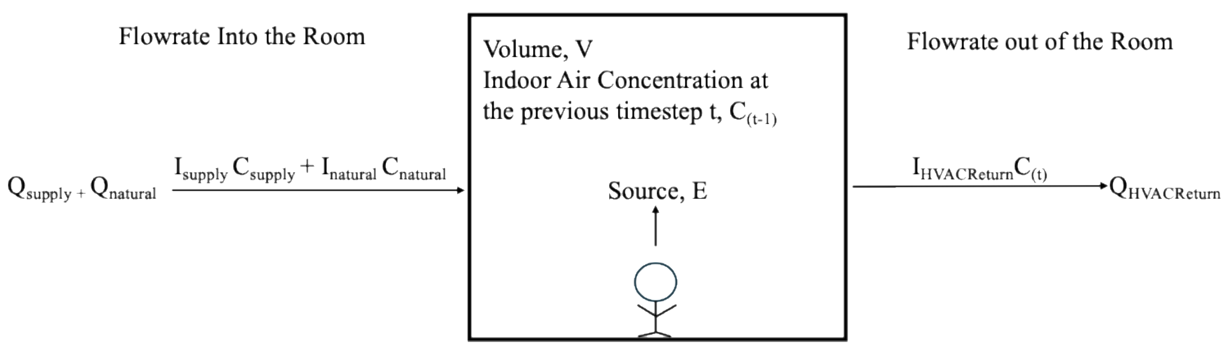

Figure 5: Box model of scenario 3 mixed mechanical and natural air exchange rate

Model Evaluation

Model evaluation was then conducted to define the relationship between CO2 concentration and occupancy patterns and to capture the effect of multiple-space ventilation systems and varying indoor environmental conditions. Using CO2 as the metric for analysis, the performance of the physics-based box models was evaluated based on the correlation coefficient to determine their deviation from the return air concentration of CO2, which represented the measured amount of CO2 leaving the classroom. This comparison focused on the correlation coefficient between the three scenarios. Additionally, to compare the CO2 concentration and occupancy across this period and determine relationship between transient occupancy and model performance during each period of occupancy (increasing and decreasing), regression models were used to determine the relationship between occupancy and CO2 concentration, assessing fit using the coefficient of determination (R^2). The r-square comparison was only done during transient occupancy, as a linear comparison could not be developed if there was minimal variability in occupancy. The r-square value analyzed was based on the models with the higher correlation coefficient, whereas the model that best represented the measured data was analyzed.

Through analyzing how the physics-based box models perform under varying environmental conditions and model assumptions, the applicability of these models for occupancy-based controls to improve energy savings, occupant comfort and indoor environmental quality was assessed. This analysis helped determine the suitability of these models, identified the conditions under which they are accurate for VAV systems, and highlighted the most influential environmental factors, including potential reasons for their significance.

RESULTS

Summary of Collected Data

In the initial stage of the data analysis the potential influence of varying environmental assumptions in models was analyzed through analysis of model performance. Table 4 and Appendix Table A.1 provide the collected datasets, length of analysis periods, temperature ranges, occupancy, ventilation rate and average and initial supply and return CO2 concentrations.

Table 4: Data summary of all periods of collected data including air exchange rates and initial levels of CO2 for the HVAC system supply and return

Door Status	Data Set Period	Time Frame (hr.)	Occupancy (range, constant/increasing/decreasing)	Natural Air Exchange Rate (ACH avg)	Mechanical Ventilation Rate (ACH avg)	Initial Return CO2 concentrations (ppm)	Initial Supply CO2 concentrations (ppm)
Both Doors Closed	1	1.3	13 ± 1, constant	0.301	0.91 ± 0.60	659	499
	2	1.0	8, constant	0.301	0.81 ± 0.46	552	487
One Door Closed	3	2	28 ± 2, constant	2.816	3.20 ± 0.28	835	448
	4	1.5	8, constant	2.816	0.84 ± 0.27	486	458
	5	0.5	27, constant	2.816	1.00 ± 0.22	688	461
Both Doors Open	6	0.9	32 ± 2, constant	2.914	1.02 ± 0.34	725	436
	7	3.3	2 ± 1, constant	2.914	0.80 ± 0.23	449	442
	8	0.5	5, constant	2.914	0.91 ± 0.57	593	496
	9	0.7	3, constant	2.914	0.82 ± 0.43	574	518
	10	1	8 ± 1, constant	2.914	0.76 ± 0.45	738	474
	11	1	0-34, increasing	2.914	0.80 ± 0.28	411	429
	12	0.8	0-17, increasing	2.914	0.91 ± 0.39	901	518
	13	0.6	0-25 increasing	2.914	0.67 ± 0.16	515	445
	14	0.6	0-27, increasing	2.914	0.76 ± 0.41	439	442
	15	0.5	28-0, decreasing	2.914	3.31 ± 0.04	744	448
	16	0.9	5-0, decreasing	2.914	0.67 ± 0.48	656	509
	17	0.5	8-2, decreasing	2.914	0.94 ± 0.29	703	490

Model Performance

Table 5 includes the correlation coefficient between the measured and modeled datasets. By using the correlation coefficient, the findings indicate the level of performance across the collected datasets. 82% of all models demonstrated strong correlation (≥ 0.9) between the modeled and measured data, as shown in Table 5. For Scenario 3, the model with natural and mechanical ventilation, was the most representative of the data, as 9 periods of collected data represented model assumptions. Scenarios 1 and 2 were less frequent, with only five and four periods of collected data that represented model assumptions. Overall findings indicated that regardless of door conditions, the number of occupants in the room, temperature, and time frame, most models performed well. Factors impacting performance were noted to be air exchange rate and high initial and average supply concentrations of CO_2 within the room.

Table 5: Correlation coefficient between the measured and modeled datasets

Door Status	Data Set Period	Correlation Coefficient[1]		
		Scenario 1, Mechanical Ventilation Dominated	**Scenario 2 Natural Ventilation Dominated**	**Scenario 3 Natural and Mechanical Ventilation**
Both Doors Closed	1	0.959	0.950	**0.966**
	2	0.952	0.928	**0.961**
One Door Closed	3	-0.330	-0.619	**0.774**
	4	0.828	0.953	**0.965**
	5	0.972	**0.990**	0.988
Both Doors Open	6	0.901	0.948	**0.973**
	7	0.360	0.918	**0.936**
	8	**0.943**	0.909	-0.922
	9	**0.973**	0.596	-0.925
	10	-0.134	**0.910**	0.564
	11	0.983	0.992	**0.993**
	12	-0.360	0.422	**0.879**
	13	0.973	**0.984**	0.942
	14	**0.990**	0.986	**0.990**
	15	0.958	**0.965**	0.842
	16	**0.889**	0.703	-0.018
	17	**0.979**	0.946	0.911

[1] *The bold values indicate the model (i.e. Scenario 1, 2 or 3) with the highest correlation coefficient*

Scenario 1: Mechanical Ventilation Dominated

In Scenario 1, the influence of the mechanical ventilation was larger than the natural ventilation of the room. Five periods of collected data represented Scenario 1 with correlation coefficients higher than Scenario 2 or 3 when comparing modeled CO_2 concentrations to measured CO_2 concentrations, *periods 8*, *9* (Figure 6a), *14* (Figure 6b), *16*, and *17* (Figure 6c). All datasets (excluding *period 16*) performed well when modeled based on the high correlation coefficients ($\geq$ 0.9) between the measured and modeled data. Correlation coefficients range from 0.943 to 0.990 (Table 5). *Period 16* had the weakest performance with the lowest correlation coefficient at 0.889. In all scenarios with high performance all doors were closed, average initial supply concentrations varied (442 ppm to 518 ppm) initial return concentrations varied (439 ppm to 703 ppm), average supply concentrations ranged from 435 to 503 ppm and average return concentrations ranged from 528 ppm to 670 ppm (Table 4 and Appendix Table A.2.). Furthermore, the amount of occupancy in each scenario varied as *period 8* and *period 9* had similar occupancy (5 people or less), *period 14* had a maximum of 27 occupants, and *period 17* had a maximum of 8 occupants. Additional information on these scenarios is also included in Appendix Table A.2.

All scenarios represented different occupancy patterns, including constant (Figure 6a), increasing (Figure 6b), and decreasing occupancy (Figure 6c), demonstrating that varying occupancy conditions showed minimal impact on accuracy of the modeling in most cases. As seen in Figure 6b and 6c, compared to the constant occupancy in Figure 6a, the box models did not track CO2 concentrations as well, with some divergence (over estimation) of CO2 concentrations particularly for the longer periods of a changing number of occupants.

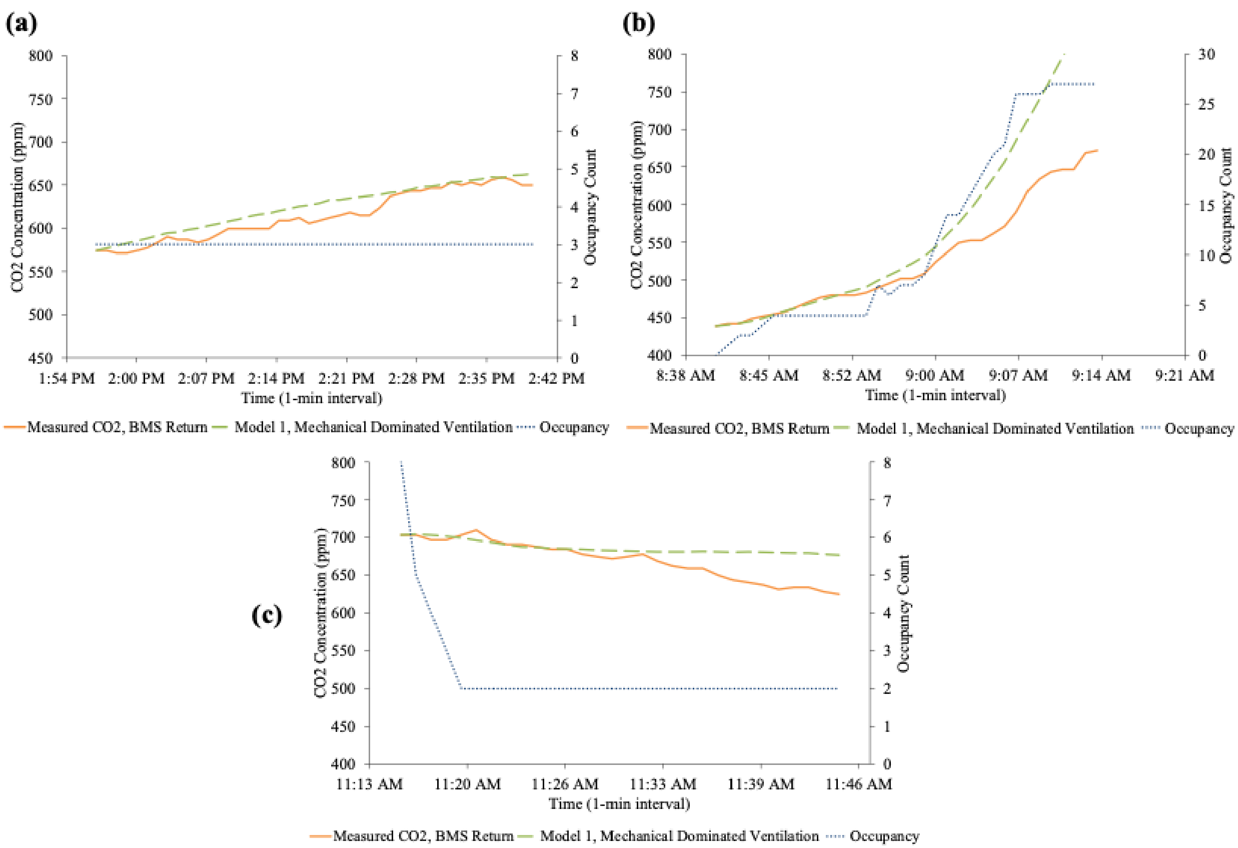

Figure 6: Graph of occupancy modeled and measured CO2 concentrations for (a) *period 9* with constant occupancy, (b) *period 14* with increasing occupancy, and (c) *period 17* with decreasing occupancy for Scenario 1 with high correlation coefficients (Mechanical Ventilation Dominated)

When comparing a dataset with weaker performance, *period 16,* as shown in Figure 7a (correlation coefficient at 0.889) to a similar dataset with stronger performance, *period 17,* (correlation coefficient at 0.979) both datasets have a similar number of occupants (range of 0 to 8) and occupant trends (decreasing occupants) (see Table 4 and Table 5). Additionally, both datasets have similar return CO2 concentrations (Appendix Table A.2). *Periods 16* and *17* have slight variations in supply CO2 concentration on average (517 ppm) than *period 17* (487 ppm), Table A.2. The main difference between *period 16* and *period 17* is the air exchange rate from the mechanical ventilation (Figure 7). *Period 16* has a lower average mechanical ventilation rate (0.67 ACH) and a higher standard deviation (0.48) indicating mechanical ventilation rate varies more, which could affect the accuracy of modeling as the natural ventilation could influence the results when mechanical ventilation was less dominate (Figure 7). Mechanical ventilation is significantly different than in *period 17* which has a higher mechanical ventilation rate (0.94) with a much lower standard deviation (0.29). As Scenario 2 is dependent on mechanical ventilation, a higher variation in ACH would logically affect the accuracy of the modeled data as it varied more. At times where mechanical air exchange rate was zero, the CO2 concentration within the room could be impacted by natural ventilation and supply CO2 concentrations, therefore impacting reliability during this dataset. This was consistent with other datasets with higher performance (*periods 8, 9 and 14)* whereas the exchange rate was generally higher (0.76 to 0.91).

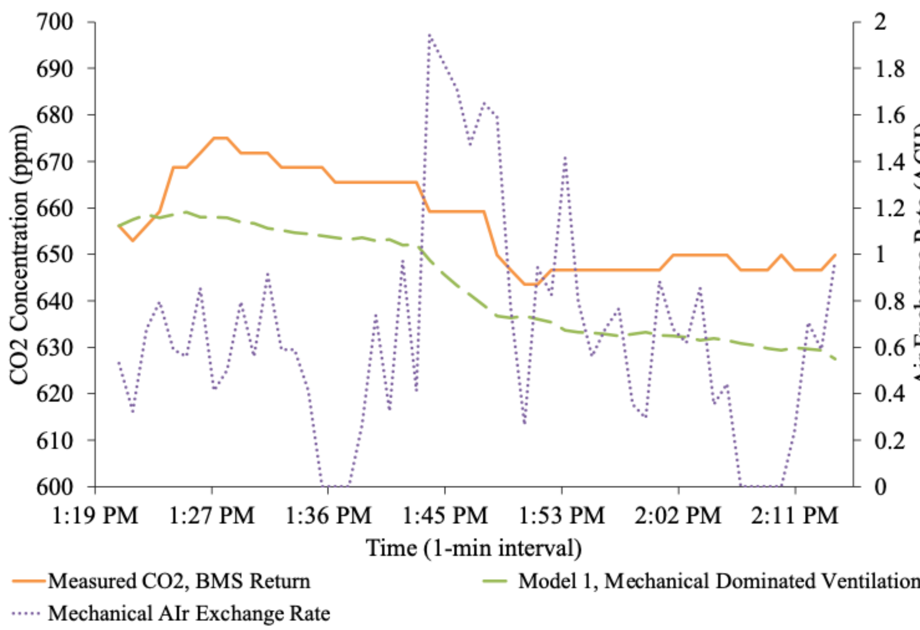

Figure 7: Mechanical and natural air exchange rate, and modeled and measured CO2 concentrations for *period 16* (poorer performance), showing higher variability of mechanical ventilation

Scenario 2: Natural Ventilation Dominated

In Scenario 2, the natural ventilation was more impactful than the mechanical ventilation of the room. Four periods of collected data (*periods 5, 10* (Figure 8c), *13* (Figure 8b) and *15* (Figure 8a)) represented this scenario with higher correlation coefficients than Scenario 1 and 3 when comparing modeled CO_2 concentrations to measured CO_2 concentrations. All modeled periods had high correlation coefficients (≥ 0.9) demonstrating a higher level of model performance (Table 5). When comparing datasets most datasets had both doors open (*period 10, 13,* and *15*), while one data set (*period 5*) had one door closed. Varying door conditions showed minimal impact on accuracy of the modeling. Additionally, the number of occupants within the room was generally higher with a maximum of 25 to 28 occupants during modeling (*period 5, 13,* and *15*), however *period 10* also displayed high correlation for this model and only had 8 ± 1 person within the room. The average initial and supply and return concentrations of CO_2 varied within the room (see Appendix Table A.2 for additional data). Major differences within this scenario were the occupancy trends, where *period 15* (Figure 8a) had decreasing occupancy, *period 13* (Figure 8b) had increasing occupancy, and *periods 5* and *period 10* (Figure 8c) had constant occupancy. This indicates that box modeling under Scenario 2 for periods of time when natural ventilation was more impactful than mechanical ventilation would be able to predict occupancy, regardless of occupancy levels or door conditions.

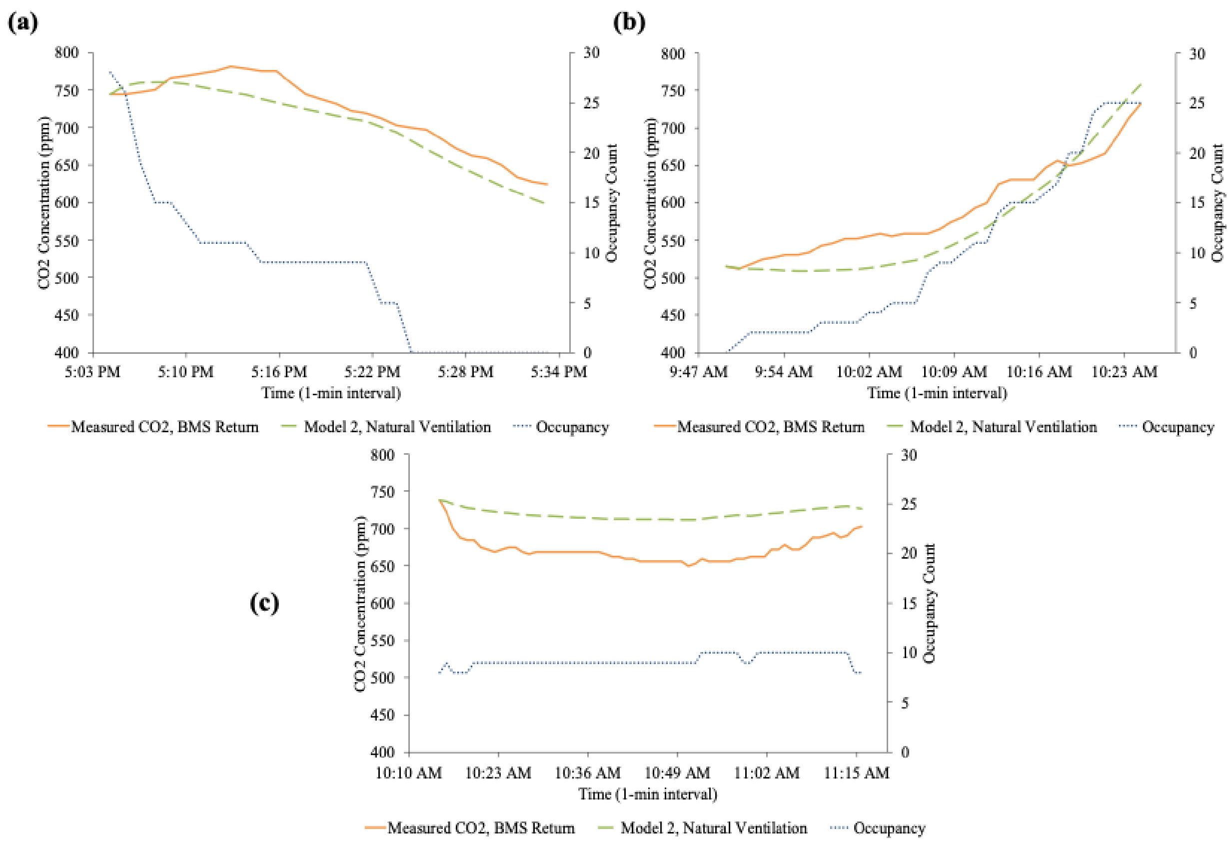

Figure 8: Occupancy modeled and measured CO2 concentrations for (a) *period 15* with decreasing occupancy, (b) *period 13* with increasing occupancy, and (a) *period 10* with constant occupancy for Scenario 2 (Natural Ventilation Dominated)

Scenario 3: Both Mechanical and Natural Ventilation

In Scenario 3, the mechanical ventilation and natural ventilation of the room were found to both be important. Nine periods of collected data represented this scenario, *periods 1-4, 6-7, 11-12*, and *14. Periods 1, 2, 4* (Figure 9a), *6, 7, 11* and *14* (Figure 9b) all have a strong correlation between the modeled and measured data (Table 5). Varying door conditions show minimal impact on the accuracy of modeling in this scenario. *Periods 1* and *2* both have two doors closed and high correlation coefficients; similarly, *period 4* (Figure 9a) has one door closed and a higher correlation coefficient; furthermore *periods 6, 7, 11*, and *14* (Figure 9b) have both doors open and high correlation coefficients (Table 4, Figure 9). Occupancy patterns and the number of occupants in the room also had minimal impact on the accuracy of modeling in this scenario as periods with high correlation coefficients have occupancy ranges from scenarios with minimum occupants 2 ± 1 (*period 7*) to the maximum number of occupants observed within the room, 34 occupants (*period*

6 and 11). Additionally, varying natural air exchange rates (0.301 ACH to 2.914 ACH) and mechanical air exchange rates (0.76 ACH to 1.02 ACH) were observed across all periods with high correlation coefficients indicating variation with ACH had little effect on model accuracy. Initial supply CO2 concentrations (436 ppm to 499 ppm) were noted to have some variation within periods with high correlation coefficients indicating limited impact on model results (Table 4; see Appendix Table A.2 for additional data). This indicates that box modeling under Scenario 3 for periods of time when natural ventilation was equivalent to mechanical ventilation influence would be able to predict occupancy, under these varied occupancy levels and door conditions.

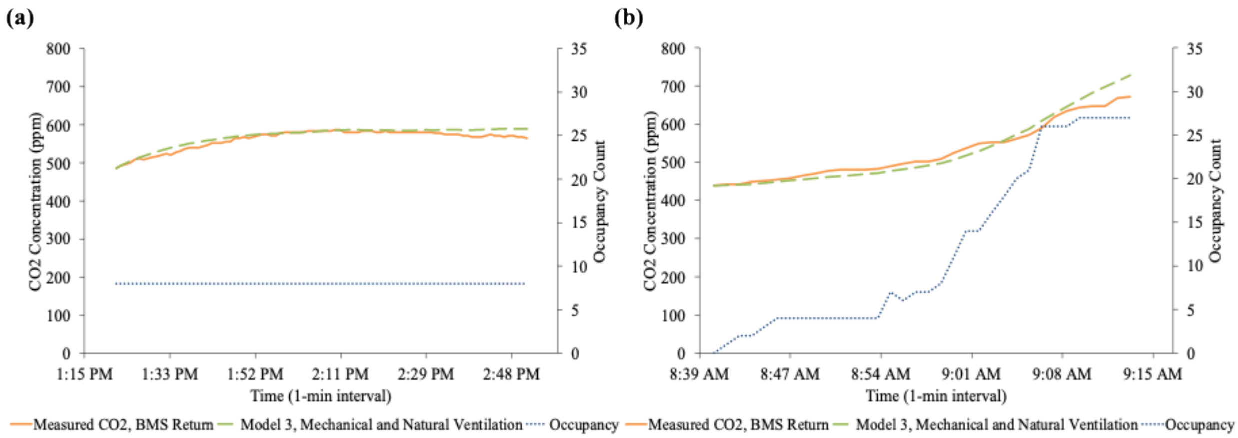

Figure 9: Occupancy and modeled and measured CO2 concentrations for (a) *period 4* with constant occupancy and one door open and (b) *period 14* with increasing occupancy and both doors open for Scenario 3 (Natural and Mechanical Ventilation)

In *periods 3* and *12* correlation coefficients showed slightly lower that other scenarios (0.774 and 0.879), as shown in Table 5. The major difference between *periods 3* and *12* (Figure 10b) with scenarios of higher accuracy (Figure 10a) in modeling was that the initial return concentrations measured within the room were higher than all other datasets within scenario 3 (835 ppm and 901 ppm), which ranged from (411 ppm to 725 ppm) . When comparing *period 12* (Figure 10b) to a period with a higher correlation coefficient of similar conditions (*period 14)*, similar conditions between these datasets where door conditions (all open) and occupancy trends (increasing). Major differences where greater occupants (27 maximum occupants) in *period 14* (Figure 10a) despite the larger number of occupants in *period 14* average measured return (582 ppm) and supply (435 ppm) concentrations of CO2 were significantly lower than *period 12* (733 ppm and 506 ppm) (see Table 4 and Appendix A.2). This indicates that for *period 12*, there may be other influential factors impacting the conditions in the room and thus the accuracy of the modeling of this dataset. This may be conditions such as occupants in the hallways or adjacent classrooms creating higher CO2 concentrations, as this modeled period increases in occupants from an initial scenario of zero occupants when CO2 concentrations are already higher. When no occupants are initially in the room, the room should be at lower CO2 concentrations. This trend is similar to *period 3* where initial CO2 concentrations within the room are high.

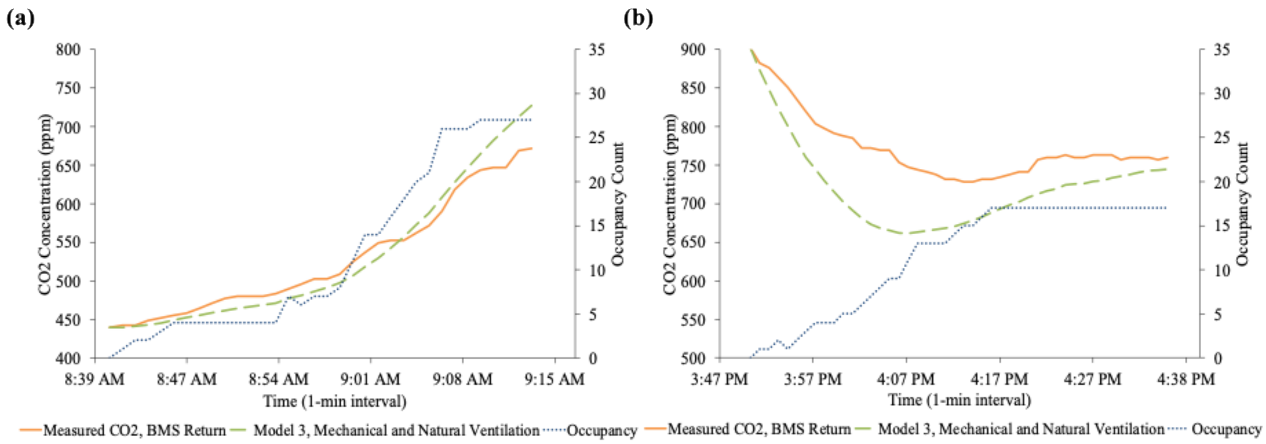

Figure 10: Graph of occupancy modeled and measured CO2 concentrations for (a) *period 14* (high performance) and (b) *period 12* (poorer performance) for Scenario 3 (Natural and Mechanical Ventilation)

Relationship Between CO2 Concentration and Occupancy

Datasets with transient occupancy were then also analyzed to examine the relationship between occupancy and CO2 concentrations during such scenarios using regression models to assessing fit using the coefficient of determination. The r-square value analyzed was based on the model with the highest correlation coefficient, whereas the model that best represented the measured data was analyzed. Overall, findings indicate that the r-squared value of a linear regression model of occupancy count based on CO2 concentrations was high during low supply and initial concentrations of CO2 into the room, low mechanical ventilation rates, and periods with steady increasing occupancy. Accuracy was lower in scenarios where CO2 concentrations were higher within the room and periods overall, along with decreasing unsteady rates of occupancy.

Table 6: R-Squared Value of Linear Regression Model of Occupancy vs. CO_2 Concentration for Transient Occupancy Periods

Data Set Period	Scenario Used Based on Correlation Coefficient	R-Squared Value[2]			
		Measured CO2 from BMS Return	**Scenario 1, Mechanical Ventilation Dominated**	**Scenario 2 Natural Ventilation Dominated**	**Scenario 3 Natural and Mechanical Ventilation**
11	3	0.87	0.82	0.88	**0.90**
12	3	0.68	0.61	0.019	**0.23**
13	2	0.95	0.96	**0.95**	0.91
14	1 or 3	0.96	**0.93**	0.94	**0.94**
15	2	0.50	0.71	**0.67**	0.88
16	1	0.002	**0.11**	0.32	0.83
17	1	0.16	**0.46**	0.35	0.43

[2] *Bolded values are analyzed based on the highest correlation coefficient among the three Models (Model 1, 2 and 3)*

For Scenario 1 (mechanical ventilation dominated) *period 11* and *14* (Figure 6b and 11a) has a high r-squared value (0.93), for Scenario 2 (natural ventilation dominated) *period 13* has a high r-squared value (0.95), and for Scenario 3 (both mechanical and natural ventilation) *period 11* and *14* have high r-squared values (0.90 and 0.94) and directly relates the increase of CO2 to occupancy (Table 6). All periods with high r-squared values are increasing in occupancy and have both doors open. Datasets demonstrate that occupancy and CO2 concentration can be directly related regardless of the number of occupants, where datasets range from 0 to 34 occupants. Other similarities within these datasets include supply CO2 concentrations, which range from initially 429 ppm to 445 ppm and are on average 432 ppm to 450 ppm (see Appendix A.2 for additional data).

Scenario 1: Mechanical Ventilation Dominated

Scenario 1 (mechanical ventilation dominated) *periods 16* and *17* both show low r-squared values (0.11 and 0.46) indicating a poor relationship between CO2 and occupancy. When comparing *period 16* and *17* with lower r-squared to a period with high predictive accuracy under Scenario 1, *period 14*, (Figure 11) similarities between these datasets include high correlation coefficients (0.965 to 0.984) between the model and measured data, demonstrating all three datasets accurately reflect measured data. However, both lower predative datasets demonstrated that when comparing just the measured CO2 in the room with occupancy there was a low r-squared value (0.002 and 0.16), indicating that the model did not cause the low predictive accuracy but a condition within the room.

Furthermore, *period 14* showed a strong correlation between CO2 levels and occupancy, with high r-squared values with both modeled and measured data (0.93 and 0.96) indicating that the model accurately predicted occupancy based on CO2 concentration. Comparing *period 14* to *16-17* datasets similar mechanical air exchange rates for *periods 16* and *14* (0.67 ± 0.48 and 0.76 ± 0.41. *Period 17* had a slightly higher mechanical air exchange rate 0.94 ± 0.29. The similar mechanical air exchange rates and similar standard deviations demonstrate that variations in mechanical air exchange rate did not influence results. Major differences for periods with low predictive accuracy compared to high predictive accuracy are the initial return (656 ppm and 703 ppm) and supply (490 ppm and 509 ppm) concentrations are higher for low predictive accuracy models compared to *period 14,* which had a high predictive accuracy model. However, *period 16-17* decreased in occupancy which makes initially high CO2 concentrations within the room more logical. However, on average, *period 14* had a lower measured return (528 ppm) and supply (435 ppm) concentration of CO2 than *periods 16* and *17* (return at 657 ppm to 670 ppm and supply at 517 ppm to 487 ppm) despite *period 14* having much greater occupancy than *periods 16* and *17*. This indicates that there could be another unaccounted for external factor influencing results when modeled as periods with lower occupancy on average are expected to have lower CO2 concentrations from the return and supply. As the ventilation system supplies a few other classrooms this could have been the source of error within this scenario.

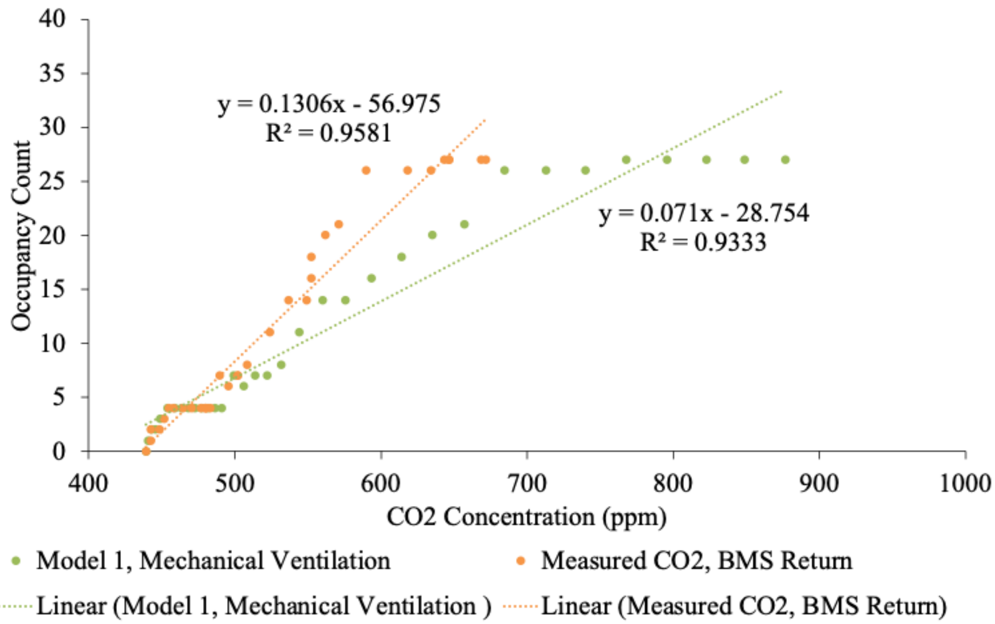

Figure 11: Predictive accuracy between occupancy and modeled CO2 concentrations for *period 14* (high predictive accuracy)

Additionally, when analyzing the rate at which people left or entered the room datasets with low r-squared values showed more drastic change in occupants. For *period 17* (Figure 12b), 6 occupants left within the first 5 minutes, however this rate drastically decreased over time and stayed at 2 occupants for the remainder of the data period. This trend was similar for *period 16* (Figure 12a) whereas at 1:21 pm, most occupants (3 occupants) left within the first 1 minute and the rest of the dataset was fairly stable with 1 or 2 occupants until 2:14 pm. However, for datasets with high r-squared, it took a longer timeframe to reach the maximum number of occupants. For example, in *period 14* (Figure 6b) it took the entire modeling period to reach max occupants. Therefore, a low r-squared result is also logical for *periods 16* and *17*, as the rate at which people enter/leave the room would affect the rate at which CO2 is released into the room and as box models assume steady state conditions.

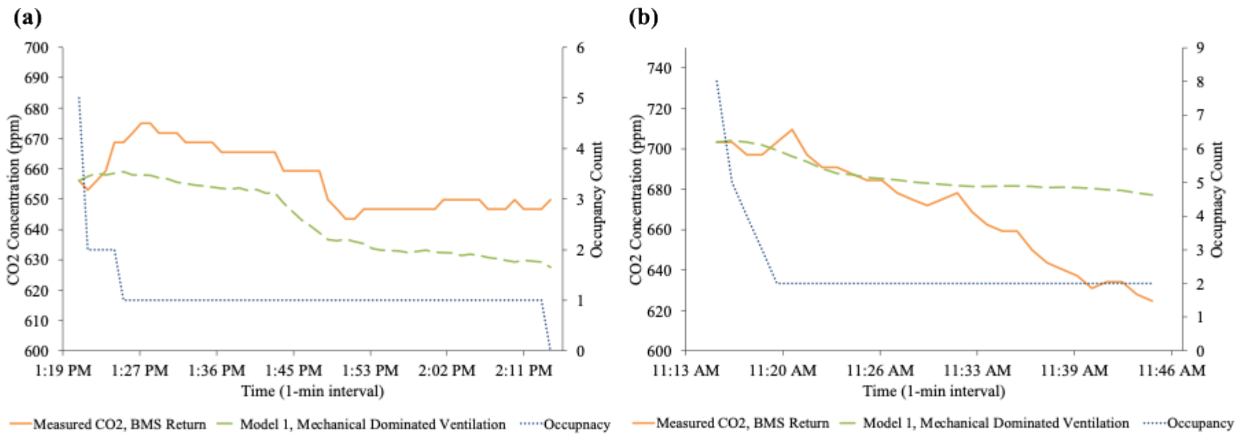

Figure 12: Modeled and measured CO2 concentrations and occupancy for (a) *period 16* (low predictive accuracy) and (b) *period 17* (low predictive accuracy) for mechanical air exchange rate dominated periods

Scenario 2: Natural Dominated Ventilation

For Scenario 2 (natural ventilation dominated) *period 15* has a low r-squared value (0.67), (Figure 12a). Similar to Scenario 1 the low predative datasets demonstrated that when comparing just the measured CO2 from the return with the relation to occupancy there was a low r-squared value (0.50), Figure 12a, indicating that the model did not cause the low predictive accuracy but a condition within the room. Supply and return CO2 concentrations where similar to other high performing datasets (see Appendix A.2 for additional data). Occupancy in this scenario was decreasing from 28 to 0 occupants over the entire time period (Figure 12b). The major difference between this dataset and others was that mechanical ACH was high, 3.31 ± 0.04 ACH. However, *period 15* did not account for mechanical ACH, was dependent on natural, as the dataset had a much higher correlation coefficient related to Scenario 2 (natural ventilation dominated). It was also noted that supply from mechanical ventilation within this room was stable at 466 ppm on average (Figure 12b). Therefore, the mechanical ventilation and supply into the room from the HVAC system should not impact this scenario. It is possible that since this model is reliant on natural ventilation, unaccounted for CO2 sources in external areas not measured (such as hallways) had higher levels of CO2. Therefore, some lag with concentrations of CO2 not decreasing as occupancy decreased would be expected, as the CO2 from the natural ventilation could possibly be high therefore CO2 concentrations would not decrease as rapidly. This can be seen on Figure 12b, where CO2 concentrations slightly decrease just not as drastic with occupancy decrease.

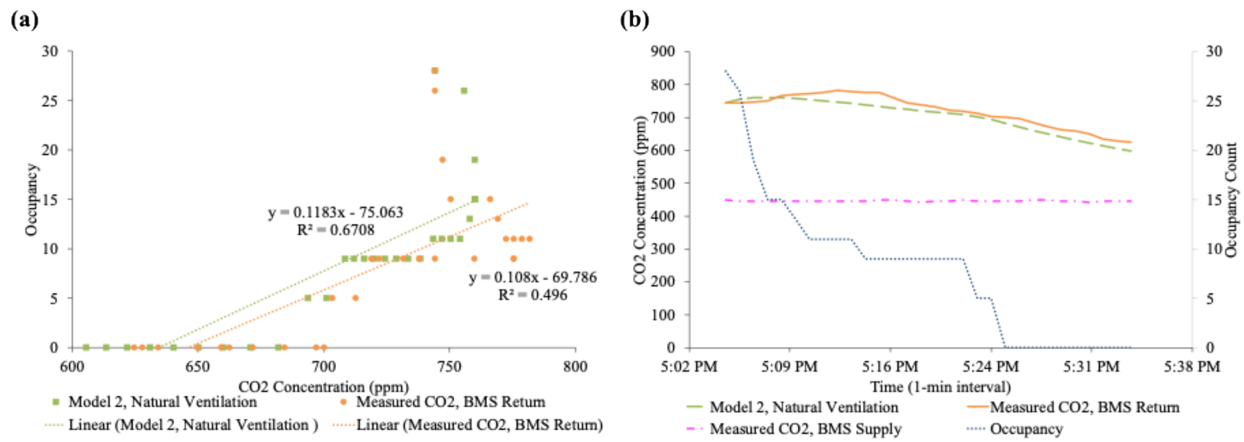

Figure 12: Comparison of modeled and measured CO2 concentrations with occupancy for (a) *period 15* (lower predictive accuracy), and comparison of modeled and measured supply and return CO2 concentrations with occupancy changes over time for (b) *period 15* (lower predictive accuracy) during periods dominated by natural air exchange rates

Scenario 3: Both Mechanical and Natural Ventilation

For Scenario 3 (both mechanical and natural air exchange) *period 12* has a low r-squared values (0.23) (Table 6). *Period 12* had a low correlation coefficient, discussed above, where inaccurate modeling could have affected the accuracy of r-squared results. This was further supported by the results from comparing the results from *period 12* Scenario 1 model and the measured BMS return r-squared values which where both higher (0.61 and 0.68) (Table 6).

CONCLUSIONS

This study aimed to define the relationship between CO_2 concentration and occupancy patterns in indoor, conditioned environment, while capturing the effect of HVAC operations, environmental, and occupancy parameters through the collection of data and utilization of physics-based box models. To accomplish this data was collected on CO_2 concentrations, tracer gas test results, door state, and HVAC operations from an academic classroom during occupied hours. Field testing using a tracer gas test was performed to determine natural air exchange rate within the classroom. Data was then modeled using physics-based models for different occupant and ventilation scenarios.

The following overall conclusions from these results are as follows:

- Most scenarios showed strong performance with a correlation coefficient of ≥ 0.9 under varying environmental conditions such as varied natural ventilation rate from door openings/closes, length of data analyzed, occupant trends (increasing, decreasing, constant), and number of occupants in the room. Periods with high rates of transient occupancy, and varying mechanical ventilation rates performed more poorly in comparison to other scenarios. Overall, this supports the use of box models for estimating occupancy trends in indoor environments for some scenarios but not others. Overall, less variable conditions resulted in better results.
- During periods of weaker performance of the correlation coefficient the air exchange rate in the room was noted to have some affect results. For example, Scenario 2, which was modeled based on the assumption of a mechanically ventilated dominated system, was impacted by natural ventilation during a period of poor performance due to the air exchange rate from the mechanical ventilation system varying greatly.
- Other periods with low correlation coefficients, specially discussed during the analysis of Scenario 3, where impacted by adjacent classrooms as the ventilation system also supplied other rooms which could have contributed unaccounted sources of CO_2 for during modeling.
- Results generally indicated that occupancy trends had a direct relationship with increasing and decreasing CO_2 concentrations in the room for models with high performance. However, only a few of the time periods across the datasets collected agreed with this. The

r-squared between occupancy count based on CO_2 concentrations was high when all doors were closed, occupancy was increasing.

- Periods showing a poor relationship between CO_2 concentrations and occupancy generally were due to non-steady state room conditions from occupants and potentially impacted by adjacent classrooms due to higher levels of CO_2 predicted than expected with lower occupants.

This study had several limitations that should be considered. One limitation that the box model assumes steady state conditions, however some conditions evaluated were not technically steady state. This could have had implications on results when comparing accuracy of predicting the relationship between CO_2 and occupancy during periods of transient occupancy, specifically during times where occupants entered or left the room rapidly or at an unsteady pace there was a low correlation between CO_2 and occupancy due to non-steady state conditions. Additionally, as noted, the rate of air flow from the hallway was not explicitly measured as compared to the HVAC air flow rate, nor was the CO_2 concentration from the hallway. In this case the CO_2 concentration then had to be assumed, which may have impacted some results. Furthermore, the data to analyze the r-squared value for transient occupancy periods was a more limited dataset and would benefit from more data. In addition, many time periods of collected data exhibited similar environmental conditions such as limited changing of temperature, similar door conditions, and non-transient occupancy. Future studies would be beneficial to evaluate box models under a broader range of varying conditions, varying spaces with different HVAC controls and periods of more transient and constant occupancy. Additionally, further work to include other aspects of VAV systems should be implemented including other variables to reflect the changing occupancy and concentrations of CO_2 in classrooms and surrounding hallways that use the same HVAC system. Additionally, the application of physics-based models could be further analyzed through the implementation of occupancy-based controls using CO_2 modeling based on its relation to occupancy trends under the conditions discussed.

The results of this study have implications that provide further insights into the applications of physics-based modeling to determine occupancy within indoor environments. Physics-based models are able to accuracy determine CO_2 concentrations within rooms. However, findings imply that the application of box modeling to determine occupancy trends for energy efficiency purposes based on CO_2 concentrations trends are only applicable during certain conditions and need to be further explored. For example, factors such as rate of occupancy change and ventilation rates were noted to influence the accuracy of results and CO_2 could not be directly related to occupancy trends. This indicates that physics-based modeling is a useful tool in modeling concentrations of CO_2 within spaces and should be further investigated with additional information and scenarios such as other aspects of VAV systems, occupancy conditions, and surrounding sources of CO_2 to assist in the outcome of understanding the applicability of this model.

REFERENCES

Automation Components, Inc. (n.d.). CO2 room sensor with temperature, setpoint & override (A/CO2-R2). https://www.kele.com/Catalog/08%20Gas_Specialty_Sensors/PDFs/A-CO2_Room_Datasheet.pdf

ARPA-E. (2017, January 18). SENSOR: Saving Energy Nationwide in Structures with Occupancy Recognition. U.S. Department of Energy. https://arpa-e.energy.gov/programs-and-initiatives/view-all-programs/sensor

ASHRAE. (2022a). ASHRAE position document on building decarbonization. https://www.ashrae.org/file%20library/about/ashrae_building_decarbonization_pd_2022.pdf

ASHRAE. (2022b). ASHRAE position document on indoor carbon dioxide. American Society of Heating, Refrigerating and Air-Conditioning Engineers. https://www.ashrae.org/file%20library/about/position%20documents/pd_indoorcarbondioxide_2022.pdf

ASTM International. (2023). Standard test method for determining air change in a single zone by means of a tracer gas dilution (Designation: E741-23). ASTM International.

Asif, A., & Zeeshan, M. (2020). Indoor temperature, relative humidity and CO2 monitoring and air exchange rates simulation utilizing system dynamics tools for naturally ventilated classrooms. Building and Environment, 180, 106980. https://doi.org/10.1016/j.buildenv.2020.106980

Bian, Y., Fu, X., Gupta, R. K., & Shi, Y. (2024). Ventilation and temperature control for energy-efficient and healthy buildings: A differentiable PDE approach. Applied Energy, 372, 123477. https://doi.org/10.1016/j.apenergy.2024.123477

Calì, D., Matthes, P., Huchtemann, K., Streblow, R., & Müller, D. (2015). CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. Building and Environment, 86, 39–49. https://doi.org/10.1016/j.buildenv.2014.12.011

Cetin, K., Davies, S., Masten, S., Lajnef, N., & Sofos, M. (2023). Simulation, challenge testing & validation of occupancy recognition & CO2 technologies (Final Scientific/Technical Report, Award No. DE-AR0001288). U.S. Department of Energy, Advanced Research Projects Agency–Energy. https://www.osti.gov/servlets/purl/1922592

Chang, F. H., Li, Y. Y., Tsai, C. Y., & Yang, C. R. (2009). Specific indoor environmental quality parameters in college computer classrooms. International Journal of Environmental Research, 3(4), 517-524. https://doi.org/10.22059/IJER.2010.67

Chu, Y., Mitra, D., Cetin, K., Lajnef, N., Altay, F., & Velipasalar, S. (2022). Development and testing of a performance evaluation methodology to assess the reliability of occupancy sensor systems in residential buildings. Energy and Buildings, 268, 112148. https://doi.org/10.1016/j.enbuild.2022.112148

Chu, Y., Mitra, D., O'Neill, Z., & Cetin, K. (2022). Influential variables impacting the reliability of building occupancy sensor systems: A systematic review and expert survey. Science and Technology for the Built Environment, 28(3), 200-220. https://doi.org/10.1080/23744731.2021.1993672

Cheng, P. L., & Li, X. (2014). Air change rate measurements using tracer gas carbon dioxide from dry ice. International Journal of Ventilation, 13(3), 235-245. https://doi.org/10.1080/14733315.2014.11684051

Dedesko, S., Stephens, B., Gilbert, J. A., & Siegel, J. A. (2015). Methods to assess human occupancy and occupant activity in hospital patient rooms. Building and Environment, 90, 136–145. https://doi.org/10.1016/j.buildenv.2015.03.029

Fan, G., Chang, H., Sang, C., Chen, Y., Ning, B., & Liu, C. (2022). Evaluating indoor carbon dioxide concentration and ventilation rate of research student offices in Chinese universities: A case study. Processes, 10(8), 1434. https://doi.org/10.3390/pr10081434

Franco, A., & Leccese, F. (2020). Measurement of CO2 concentration for occupancy estimation in educational buildings with energy efficiency purposes. Journal of Building Engineering, 32, 101714. https://doi.org/10.1016/j.jobe.2020.101714

Gu, J., Yang, B., Brauer, M., & Zhang, K. M. (2021). Enhancing the evaluation and interpretability of data-driven air quality models. Atmospheric Environment, 246, 118125. https://doi.org/10.1016/j.atmosenv.2020.118125

Huang, Q., Syndicus, M., Frisch, J., & van Treeck, C. (2024). Spatial features of CO2 for occupancy detection in a naturally ventilated school building. Indoor Environments, 1(3), 100018. https://doi.org/10.1016/j.indenv.2024.100018

Kong, M., Dong, B., Zhang, R., & O'Neill, Z. (2022). HVAC energy savings, thermal comfort, and air quality for occupant-centric control through a side-by-side experimental study. Applied Energy, 306(Part A), 117987. https://doi.org/10.1016/j.apenergy.2021.117987

Kula, B., Mitra, D., Chu, Y., Cetin, K., Gallagher, R., & Banerji, S. (2023). Laboratory testing methods to evaluate the reliability of occupancy sensors for commercial building applications. Building and Environment, 240, 110457. https://doi.org/10.1016/j.buildenv.2023.110457

Lawrence, T. M., & Braun, J. E. (2007). A methodology for estimating occupant CO2 source generation rates from measurements in small commercial buildings. Building and Environment, 42(2), 623–639. https://doi.org/10.1016/j.buildenv.2004.10.021

Li, H., Li, X., & Qi, M. (2014). Field testing of natural ventilation in college student dormitories (Beijing, China). Building and Environment, 78, 36–43. https://doi.org/10.1016/j.buildenv.2014.04.009

Li, T., Liu, X., Li, G., Wang, X., Ma, J., Xu, C., & Mao, Q. (2024). A systematic review and comprehensive analysis of building occupancy prediction. Renewable and Sustainable Energy Reviews, 193, 114284. https://doi.org/10.1016/j.rser.2024.114284

Mannan, M., & Al-Ghamdi, S. G. (2021). Indoor air quality in buildings: A comprehensive review on the factors influencing air pollution in residential and commercial structures. International Journal of Environmental Research and Public Health, 18(6), 3276. https://doi.org/10.3390/ijerph18063276

Miao, S., Gangolells, M., & Tejedor, B. (2023). Data-driven model for predicting indoor air quality and thermal comfort levels in naturally ventilated educational buildings using easily accessible data for schools. Journal of Building Engineering, 80, 108001. https://doi.org/10.1016/j.jobe.2023.108001

Mitra, D., Koupaei, D. M., & Cetin, K. (2022). Occupancy data sensing, collection, and modeling for residential buildings. In Renewable energy for buildings (pp. 103–121). Springer. https://doi.org/10.1007/978-3-031-08732-5_6

Occupational Safety and Health Administration. (n.d.). Carbon dioxide in workplace atmospheres (ID-172). U.S. Department of Labor. Retrieved from https://www.osha.gov/sites/default/files/methods/osha-id172.pdf

Raspberry Pi Foundation. (n.d.-a). Camera modules. Raspberry Pi. Retrieved January 22, 2025, from https://www.raspberrypi.com/documentation/accessories/camera.html

Raspberry Pi Foundation. (n.d.-b). Raspberry Pi 4 Model B specifications. Raspberry Pi. Retrieved January 22, 2025, from https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/

SAF Tehnika JSC. (2025). Aranet4 PRO: Wireless, portable device for measuring air quality (Version 3.4). https://aranet.com/attachment/613/Aranet_Datasheet_TDSPC003_Aranet4_PRO.pdf

Sun, Z., Wang, S., & Ma, Z. (2011). In-situ implementation and validation of a CO2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. Building and Environment, 46(1), 124–133. https://doi.org/10.1016/j.buildenv.2010.07.008

U.S. Energy Information Administration. (2024, December). Monthly energy review: Table 2.1a energy consumption: Residential, commercial, and industrial sectors. Retrieved January 21, 2025, from https://www.eia.gov/totalenergy/data/monthly/pdf/sec2_4.pdf

Vaisala. (2023). GMP252 carbon dioxide probe: For ppm-level measurements. Vaisala. https://docs.vaisala.com/v/u/B211567EN-G/en-US

Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H. B., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy and Buildings, 107, 264–278. https://doi.org/10.1016/j.enbuild.2015.08.032

Yang, J., Pantazaras, A., Chaturvedi, K. A., Chandran, A. K., Santamouris, M., Lee, S. E., & Tham, K. W. (2018). Comparison of different occupancy counting methods for single system-single zone applications. Energy and Buildings, 172, 221-234. https://doi.org/10.1016/j.enbuild.2018.04.051

Yang, L., Wang, X., Li, M., Zhou, X., Liu, S., Zhang, H., Arens, E., & Zhai, Y. (2020). Carbon dioxide generation rates of different age and gender under various activity levels. *Building and Environment, 186*, 107317. https://doi.org/10.1016/j.buildenv.2020.107317.

Zuraimi, M. S., Pantazaras, A., Chaturvedi, K. A., Yang, J. J., Tham, K. W., & Lee, S. E. (2017). Predicting occupancy counts using physical and statistical CO2-based modeling methodologies. Building and Environment, 123, 517–528. https://doi.org/10.1016/j.buildenv.2017.07.027

APPENDIX

Table A.1: Field test natural air exchange rate for both doors open, both doors closed, and one door closed and one door open

Sensor Name	Both Doors Open (ACH)	Right Door Closed (ACH)	Left Door Closed (ACH)	Both Doors Closed (ACH)
Sensor 2	0.321 ± 0.017	2.770 ± 0.168	2.830 ± 0.048	3.292 ± 0.095
Sensor 3	0.001 ± 0.161	2.576 ± 0.099	2.432 ± 0.028	0.182 ± 0.149
Sensor 4	0.307 ± 0.000	2.686 ± 0.219	3.225 ± 0.082	2.853 ± 0.090
Sensor 5	0.247 ± 0.019	2.844 ± 0.068	2.558 ± 0.038	3.082 ± 0.126
Sensor 6	0.208 ± 0.049	2.947 ± 0.119	2.749 ±0.030	2.014 ± 0.079
Sensor 7	0.301 ± 0.023	2.775 ± 0.144	2.840 ± 0.037	1.017 ± 0.151
Average	**0.3010 ± 0.011**	**2.816 ± 0.122**	**2.796 ± 0.017**	**2.914 ± 0.042**
BMS (ACH, avg)	**0**	**0.11**	**0.17**	**0**

Table A.2: Data summary of all periods of collected data including average CO2 concentrations for both supply and return collected from the BMS and temperature range (C)

Door Status	Data Set Period	Return CO2 (ppm) (avg)	Supply CO2 (ppm) (avg)	Temperature range (C)
Both Doors Closed	1	800	490	22 to 22
	2	650	492	21 to 22
One Door Closed	3	799	449	23 to 23
	4	562	458	22 to 23
	5	806	468	21 to 22
Both Doors Open	6	811	445	22 to 23
	7	478	441	22 to 22
	8	632	503	22 to 22
	9	614	510	21 to 22
	10	672	482	22 to 22
	11	582	432	21 to 23
	12	733	506	22 to 22
	13	589	450	21 to 22
	14	528	435	21 to 21
	15	721	446	23 to 23
	16	657	517	22 to 22
	17	670	487	22 to 22