THESIS

FIELD-BASED APPROACHES TO CHARACTERIZING LONG-TERM INDOOR ENVIRONMENTAL QUALITY IN HOMES

Submitted by

Andrew Purgiel

Department of Civil & Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2022

Master's Committee:

Advisor: Ellison Carter

Jens Blotevogel
Tami Bond

ABSTRACT

FIELD-BASED APPROACHES TO CHARACTERIZING LONG-TERM

INDOOR ENVIRONMENTAL QUALITY IN HOMES

The overall goal of this research was to evaluate the performance of energy and indoor environmental quality (IEQ) metrics for future use in impact evaluations of residential energy efficiency upgrades. Analysis focused on the temporal representativeness, spatial representativeness, and spatial specificity of indicators, with an aim to answer the questions:

1. To what extent do shorter-duration measurements of indoor environmental quality (IEQ) indicators characterize long-term trends observed in daily IEQ conditions within a sampling location?
2. To what extent do single-zone measurements of IEQ indicators in a home characterize multi-zone IEQ conditions within the home?
3. What specific information is gained by measuring IEQ in more than one room within a home?

The spatial and temporal patterning of indoor environmental quality (IEQ) metrics were observed using commercial-grade IEQ sensors in the living room, bedroom, kitchen, garage, and outdoors for 15 owner- and renter-occupied single-family homes in the City of Fort Collins, Colorado. Indicators of IEQ, including: fine particulate matter ($PM_{2.5}$), a measure of total volatile organic compounds (TVOC), carbon dioxide (CO_2), temperature (T), relative humidity (RH), light, noise, and energy use were monitored continuously in each home for six to ten months.

The number of hours for which valid IEQ sensor data were recorded from indoor locations (bedroom, kitchen, and living rooms) ranged from 3,248 hours (136 days) to 7,507 hours (315 days), with a median of 6,589 hours (275 days) across all homes. Time weighted hourly average values of indoor concentrations, pooled across all homes, were calculated over the entire study period for $PM_{2.5}$ (mean: 8.2 μg/m^3, standard deviation: 27.0 μg/m^3, coefficient of variation: 3.27), TVOC (mean: 340 ppb, standard deviation: 377 ppb, coefficient of variation: 1.11), and CO_2 (mean: 749 ppm, standard deviation: 364 ppm, coefficient of variation: 0.49).

Seasons were defined by daily participant heating and cooling behaviors. These behaviors were measured using one-minute resolution energy use data from heating (e.g., furnace) and cooling (e.g., air conditioning) devices within each home. Overall, median $PM_{2.5}$, TVOC, and CO_2 concentrations were lower in the heating season than in the cooling and shoulder seasons. Ranges of indoor $PM_{2.5}$, TVOC, and CO_2 concentrations were comparable between seasons.

Hour-of-day average trends of $PM_{2.5}$ suggested cooking activities in the kitchen were significant sources of $PM_{2.5}$ in most homes. Average $PM_{2.5}$ concentrations increased at similar hours of the day between living rooms, kitchens, and bedrooms. Bedroom and living room evening peaks (around 6pm) yielded lower $PM_{2.5}$ concentrations, on average, compared to kitchen evening peaks. Hour-of-day average TVOC trends in kitchens and living rooms displayed evening peaks that were likely attributed to garage sources or increased indoor participant activity (i.e., cooking and cleaning). Correlations between $PM_{2.5}$ hourly concentrations recorded in the garage and those recorded in indoor rooms were observed to vary with a predictable pattern throughout the day. If future studies investigated drivers and determinants of this garage-to-indoor relationship, we may expect to discover more on the mechanisms of infiltration of $PM_{2.5}$ and other pollutants from garages and outdoor areas into living spaces.

The extent to which in-home hourly $PM_{2.5}$, TVOC, and CO_2 samples (sampling periods ranging from one day to fourteen days in all three seasons) represented IEQ conditions over a long-term (six- to ten-month) period was evaluated. This evaluation was performed using a measure defined as *time-structured temporal representativeness*. This measure quantifies how well the average hour-of-day structure for a long-term monitoring period is characterized by data from a shorter sampling period (i.e., how representative the shorter sampling period is). A threshold value was defined to identify when a sample is considered representative. Temporal representativeness of samples increased with sample length. Depending on the season, 80 to 91% of three-day $PM_{2.5}$ samples and three-day TVOC samples were considered representative. Representativeness of $PM_{2.5}$ and CO_2 samples varied by season. Analysis suggested practitioners sampling IEQ indicators can be confident in the time of day at which $PM_{2.5}$ or TVOC peaks occur on a "typical" day, based on three-day samples; CO_2 samples may require longer lengths. Even if resources are only available to sample for one day, our analysis suggested the time structure of a $PM_{2.5}$ sample (i.e., the hour(s) when concentration peaks during the day) has a high likelihood of being representative of a "typical" day; however, this likelihood may vary depending on sampling season.

The measure of spatial representativeness of IEQ samples was defined in the current study to evaluate how well data gathered from a sampling location (a room) captures trends and magnitudes that characterize average conditions within the larger location of interest (the home). The complementary measure of spatial specificity was used to evaluate how well data recorded in a room captures trends and magnitudes that are not captured by samples recorded in other sampling locations within the home (i.e., how specific or unique the room's data are). In most homes, $PM_{2.5}$, TVOC, and CO_2 data recorded in bedrooms were the most specific of all three indoor rooms (bedroom, living room, and kitchen), but the least representative. These results suggest that if practitioners are aiming to observe the full range of IEQ variability between living spaces, and they are only able to install IEQ sensors in two rooms within a home, the bedroom should be one of the rooms sampled. However, the data gathered from the bedroom may only be applicable if conditions within the bedroom are of interest. Relationships between room, spatial representativeness, spatial specificity, and other variables, such as distance between rooms and HVAC structure, could be explored to discover why between-room variability is higher in certain homes compared to others. Understanding these relationships would help practitioners estimate how many sensors are needed within a home to characterize IEQ conditions within living areas, given building characteristics and the focus of the sampling campaign.

This study was conducted in partnership with the Epic Homes program, the purpose of which is to improve the energy efficiency of Fort Collins homes (while also improving the health and well-being of residents) by offering technical and financial assistance for home energy efficiency upgrades. Findings have implications for those aiming to develop best practices when taking short samples of IEQ indicators in homes, whether they be energy efficiency practitioners determining the impacts of residential upgrades or researchers considering IEQ impacts on occupant health.

ACKNOWLEDGEMENTS

I'd like to say thank you to my advisor, Dr. Ellison Carter, for guiding me though the research process and continuously offering support throughout the development of this thesis. I'd also like to thank my other committee members, Dr. Tami Bond and Dr. Jens Blotevogel. I've learned much from both of them on topics related to research, professional development, and future goals.

I also must thank Dr. Nicholas Good and all those in Dr. Carter's research group who helped with the collection and interpretation of data: Kelsey Walker, Collin Brehmer, Troy Montague, Florencio Guzman Diaz, and Dr. Oluwatobi Oke. I received fantastic advice and ideas for analysis through this group. The data collected for this thesis would not have been nearly as complete without constant home visits and damage control for malfunctioning sensors.

Thank you to those in Dr. Bond's research group: Sagar Rathod, Paige Ambord, and John Flynn. I gained a wealth of advice on data presentation and interpretation through our meetings. I also engaged in unexpected, but well needed conversations of life plans, past experiences, and hot yoga.

Thank you as well, everyone at Fort Collins Utilities involved in the Epic Homes program. The study would not have been possible without your support. I learned so much about the field of indoor environmental quality by working alongside the program.

Thank you, Dr. Julia Sharp with the Franklin A. Graybill Statistical Laboratory, for helping me unravel the mysteries of statistical measures related to relative entropy.

Thank you, Dr. Sybil Sharvelle and Lucas Loetscher for all the support while providing me my first research opportunity here at Colorado State University. I enjoyed the opportunity to delve into hands-on chemistry work, especially with topics related to the field of wastewater treatment.

Thank you to all my family, friends, and roommates who helped me enjoy my past two years at Colorado State University, especially during the days researching while in COVID-19 isolation.

Lastly, I'd like to acknowledge this work was supported by the City of Fort Collins, a 2018 winner of the Bloomberg Philanthropies U.S. Mayors Challenge, and by Harvard University and the JPB Foundation in support of their Environmental Health Fellows (Carter)."

TABLE OF CONTENTS

1. Introduction

Exposure to pollutants in outdoor and indoor air, including gases and aerosols, contribute to national burden of disease (Szigeti et al., 2017). Lack of thermal comfort, generally measured by appropriate indoor temperature and relative humidity, has also emerged as a concern for residential physical and mental health (E4 the Future, 2016; Hayes et al., 2020; International Energy Agency, 2014; Vermont Energy Investment Corporation, 2019; Wilson et al., 2016; World Health Organization, 2018). Indicators of indoor air quality (IAQ) and thermal comfort together comprise indicators of indoor environmental quality (IEQ). IEQ-related variables and their associated health effects are important to study, as people spend a significant amount of time indoors. Gathering data from 1992 to 1994 from U.S. respondents (n= 9,386), the National Human Activity Pattern Survey found on average, 87% of people's time was spent in enclosed buildings; most of this (69% of overall time) was spent in residences (Klepeis et al., 2001). Similarly, in more recent data collected in 2010-2011 from 5,011 Canadian respondents (infants and children were over-sampled because they are a common focus in risk assessment), the Canadian Human Activity Pattern Survey 2 found an average of 89% of respondents' time was spent indoors, with 70% of total time spent at home (Matz et al., 2014).

Opportunities to make changes to homes, once built, are limited. Retrofits and upgrades that focus on improving energy efficiency are one potential inroad to homes. Homes in the U.S. account for approximately 20% of national energy consumption, which motivates programming to increase residential energy efficiency. At the same time, lowering energy costs for low-income households has been a long-standing goal at federal, state, and local levels through programs such as the Weatherization Assistance Program (WAP) (U.S. Department of Energy, 2022). Reducing greenhouse emissions generated from the heating and cooling of homes is one of the drivers for these programs.

Another driver is reducing energy burden (the amount of money going toward energy bills) for low-income residents (Tonn et al., 2018). High energy burden is an issue that is prevalent nationwide, including in Colorado. Fisher Sheehan & Colton (2020) provided analysis on energy burden for U.S. residents. Their analysis was based on American Community Survey data from the U.S. Census Bureau and a county-by-county model that accounts for energy use, housing/residential characteristics, and climate. Fisher Sheehan & Colton (2020) estimated that Colorado residents living at or below the poverty line spend 12 to 22% of their monthly income on home energy bills.

Indoor environmental quality (IEQ) and home energy use impact the social, physical, and financial well-being of residents. Increasingly, relationships and interactions between IEQ and home energy use are documented – e.g., inadequate IEQ, high energy use, and poor resident health – are becoming more of a concern (Becchio et al., 2018; Hayes et al., 2020). There is a growing demand to characterize and quantify a broader set of impacts that may arise from energy efficiency upgrades (Broderick et al., 2017). Original metrics used to evaluate WAP-funded programs through the 1990 National Evaluation of the Weatherization Assistance Program focused on energy savings for occupants and cost effectiveness for organizations administering the funding (Oak Ridge National Laboratory, 1990). There have been many follow-up retrospective evaluation studies sponsored by the Department of Energy since then; however, indoor environmental quality, exposure, and health-related impacts of weatherization programs are notably absent from these extensive evaluation efforts (U.S. Department of Energy & Oak Ridge National Laboratory, 2010). One barrier to monitoring and evaluating a more diverse set of non-energy impacts is the limited guidance on how to measure indoor environmental quality and energy use in homes to evaluate the "health" of these homes (Wei et al., 2020). Studies also often overlook, or are not able to adequately track, the behaviors of those living within the home. These resident behaviors can impact a home's IEQ and energy use.

The development of a rating system or framework for evaluation and relative comparison that incorporates IEQ and energy use metrics (related to both home characteristics and resident behavior) would be beneficial for practitioners who are looking to evaluate and improve homes for the sake of resident health. Establishing such a framework requires identification and definition of metrics to be used within the framework. These metrics must be resource-efficient and must effectively capture meaningful variability in IEQ and energy use data for the purpose of placing a home on an index or within a scoring framework. To this end, we conducted a year-long study in single-family homes in Fort Collins, Colorado to evaluate the performance of energy and IEQ metrics for future use in impact evaluations of residential energy efficiency upgrades.

2. Literature Review

2.1. Evaluation of Home Energy-Efficiency Upgrades

A framework for rating non-energy impacts associated with energy efficiency upgrades could help standardize evaluation of a broad spectrum of upgrades that are commonly installed in homes across the United States and other countries. Examples of such upgrades are listed in Table 1.

Table 1: List of energy efficiency upgrades that can be evaluated with frameworks using the metrics measured in this study (Wilson et al., 2016)[1]

Energy Efficiency Upgrade Type
• Improved/additional insulation[1]
• Sealing of cracks in building envelope[1]
• Improved house ventilation system[1]
• Improved cooking system[1]
• Improved heating system[1]
• Improved cooling system
• Installation of energy efficient windows
• Improved light fixtures

Such a framework could be comprehensive by including non-energy impacts, or benefits (often called co-benefits in air quality literature) of energy efficiency upgrades. However, both the scientific research enterprise and the practice of residential energy efficiency upgrades lack frameworks that include indoor environmental quality (IEQ) indicators together with measures of energy efficiency performance.

Scientific studies that evaluate energy efficiency upgrades using energy use or impacts on cost savings are prevalent in literature. In their review on the subject, Fregonara & Pattono (2018) found over 100 studies that evaluated building projects with life cycle-based approaches, most including energy, environmental, and economic effects of renovation projects. Pombo et al. (2016) reviewed 42 studies that considered sustainability indicators of building renovations. Within the 42 studies, the authors of the review noted a focus on multi-criteria assessments that had the goal of improving energy efficiency by balancing cost and quality. Pombo et al. (2016) suggested a life-cycle approach be taken for effective assessment of upgrades. It was also noted that none of the reviewed studies had included social impacts of upgrades – i.e., consequences of upgrades on “social endpoints,” such as the wellbeing of residents, had not been considered.

Evaluating the health impacts of energy efficiency upgrades by considering upstream risk factors for health, such as indicators of IEQ, is a newer area of research than evaluating the energy use or cost impacts. Patino et al. (2018) reviewed the literature and found 49 studies that focused on changes in IEQ indicators associated with energy efficiency upgrades in public housing. These authors noted no attempts to create a framework for evaluation. The U.S. Department of Energy (Wilson et al., 2016), sponsored a review of the scientific literature on positive impacts on resident health that can occur from energy efficiency upgrades. The systematic review focused on health-care utilization data to monetize impacts of upgrades. The authors also provided an overview of IEQ metrics commonly used in studies that look at the health impacts of healthy home interventions. Commonly used IEQ metrics included physical and chemical contaminants (e.g., particulate matter, volatile organic compounds, nitrogen dioxide, carbon monoxide, radon, and carbon dioxide), thermal comfort parameters (e.g., temperature and relative humidity), and non-airborne contaminants (e.g., indoor allergens, mold, musty smells, dampness, and evidence of pests). The authors concluded their review with a call for more research on the impacts of upgrades on specific health outcomes to: (a) gain more buy-in from health care professionals for energy efficiency programs and (b) accelerate greater integration between energy and health-motivated initiatives, policies, and programs. For example, the Energy-Plus-Health Playbook (Vermont Energy Investment Corporation, 2019) is a report created by a coalition of energy-efficiency and health-related practitioners. This coalition called for programs that would build partnerships between professionals responsible for installation of energy efficiency upgrades and healthcare providers. The Energy-Plus-Health Playbook suggested many metrics for the evaluations of homes with a focus on resident health, including the cost and timelines of energy efficiency programs, pre- and post-intervention measurements of IEQ indicators, and surveys of health, well-being, and customer satisfaction. An anticipated benefit of energy and health partnerships, as described in the 2019 report, would be linkage between health data, home energy efficiency upgrades, and (potentially) avoided medical costs. Partnerships such as these could, ultimately, provide the evidence base needed to quantify and monetize hypothesized health and well-being benefits of home energy efficiency upgrades. This report called these health impacts Non-Energy Impacts (NEIs) and asserted that upgrades in the future could be evaluated by summing household annual value and societal value. A research report produced by ACEEE (Hayes et al., 2020) aimed to monetize the health impacts of home energy efficiency upgrades on residents. The report considered possible health benefits of energy efficiency upgrades related to four major health threats: asthma, cold-related thermal stress, heat-related thermal stress, and trip-an-fall injuries. A general formula was developed to monetize health impacts of upgrades based on the possible adverse healthcare costs avoided. This paper suggested the resulting monetized values could be used to provide information that decision makers in both energy and health sectors could use to evaluate these upgrades. The authors also stated that the monetized values could attract additional funding for residential energy efficiency upgrade programs.

Increasingly, researchers and practitioners are directing effort toward the development of frameworks that incorporate energy and non-energy impacts into decision-making tools for building energy efficiency upgrades. In this field, researchers have attempted to create algorithms that will help energy-efficiency practitioners and homeowners decide upon optimal upgrades for a home depending on multiple factors. Both Turner et al. (2013) and Das et al. (2013) determined the optimal level of ventilation within modeled homes by weighting energy savings and resident health/IAQ impacts. The aim of their studies were to create algorithms that could be used to evaluate ventilation-related energy efficiency upgrade options for a home before installment. Studies by Touceda et al. (2018), Ortiz et al. (2019), Ezratty et al. (2018), and Underhill et al. (2020) all modeled the impacts of upgrades on specific housing types, combining energy/economic impacts with health impacts. Health impacts were estimated for modeled changes in IEQ indicators and were summed with economic impacts after being monetized. An additional study by Underhill et al. (2018) added a key component to their model that others had lacked: the behavior of residents (e.g., smoking, cooking methods, and window opening).

Only considering building structure in models is likely ignoring key factors; building characteristics alone have been shown to be ineffective at predicting dependent IEQ indicators (Clements et al., 2019). Therefore, inclusion of empirical monitoring of IEQ indicators would likely add value in a building energy efficiency evaluation framework. Residential behavior has been shown to impact IEQ indicators (Hollnagel, 2014). However, distributions of, and heterogeneity in, resident behaviors are often absent in decision making tools used to select among home energy efficiency upgrade choices (Underhill et al., 2020). This absence is because resident behaviors are challenging to measure and/or model. Overlooking residential behaviors likely contributes to implementation and performance gaps between anticipated and realized benefits of home energy efficiency upgrades.

Studies that have attempted to create a framework to evaluate energy efficiency upgrades using pre- and post- measurements of energy and IEQ indicators are few. In a 2013 study, Xu et al. developed a framework for energy efficiency retrofits of hotel buildings in China to evaluate programs in which contractors assess buildings and offer retrofit projects at discounted rates. The authors of this study developed the framework by modifying a framework typically used to evaluate organizations – the "EFQM Excellence Model." This strategy of framework development resulted in a heavy focus on the management of the project, and less of the pre- and post- retrofit performance of the building with respect to IEQ indicators. Two of the eleven proposed performance indicators included in Xu et al.'s study are "Health and Safety" and "Environmental Loading," though these are not considered in detail. This study seemed to be a promising start to a framework or rating system that could be used to evaluate residential upgrades; social, environmental, and economic sustainability of projects were considered, and the framework stressed the importance of including building operations as well as building characteristics.

In another study, Wei et al. (2020) aimed to develop "a measurement protocol and a systematic method for rating IEQ that could also be used to estimate any non-energy benefits associated with improved IEQ that can add financial value." The authors focused on offices and hotels in their survey of literature and building certification programs. The study evaluated and compared 55 current "certification schemes" used in the industry of building evaluation that had at least a partial focus on IEQ parameters. Wei et al. selected fourteen of these schemes for a detailed review, as these fourteen schemes defined criteria for a non-residential building. These fourteen schemes included well-known building certifications – e.g., Leadership in Energy and Environmental Design (LEED) (U.S. Green Building Council, 2021) and Building Research Establishment Environmental Assessment Methodology (BREEAM) (Building Research Establishment Ltd, 2022) – that consider various building aspects such as energy, water, and materials, in addition to IEQ indicators. Wei et al. noted that three of the building certifications considered in the paper, OsmoZ (Certivéa, 2022), BES – i.e., Bienestar en Espacios Sostenibles – (Institue Valencià de l'Edificació, 2022), and WELL (International WELL Building Institute, 2022), focused primarily on the health and well-being of those using the building, or the quality of life allowed within the building. The definitions of health, well-being, and quality of life are often broadly defined. Wei et al discuss in their supplemental material how the definition of health can be broad; however, the definitions Wei et al. select from three sources (World Health Organization, Merriam Webster Dictionary, and an unidentified medical source) converge well. While Wei et al. do not define well-being and quality of life, all three topics (health, well-being, and quality of life) are included in the authors' definitions of IEQ. These multiple IEQ definitions converged on the conclusion that IEQ indicators can have an impact on occupant health, well-being, and quality of life. This convergence suggests that scientist and practitioners in the field of building evaluation are coming to an agreement on the importance of IEQ on occupant health and the importance of including IEQ in building evaluation frameworks. Wei et al. themselves did not elaborate further on the WELL, OsmoZ, and BES schemes. The WELL scheme seems heavily centered around IEQ indicators. English translations of the OsmoZ and BES schemes were not available for further detail in our study.

In the aforementioned study (Wei et al., 2020), the authors sorted all the IEQ indicators used across the 55 considered schemes into four categories: (i) thermal environment (including temperature set points inside the building); (ii) IAQ parameters (including pollutants such as $PM_{2.5}$ and VOCs); (iii) acoustic environment (mainly ambient noise); and (iv) and visual environment (including illuminance level and measures of daylight). The number of “credits” assigned to each category by all schemes combined was used to estimate the relative importance given to each IEQ category.

The authors then used a second method to estimate the relative importance of the four considered IEQ categories. This method considered the rating systems or frameworks developed in nine peer reviewed studies. Each of these nine frameworks included IEQ indicators that belonged to the four considered IEQ categories. Each framework also assigned a weight to each IEQ category and aggregated the weights to assign a single score to a building. The studies that had tested their developed rating system or framework on existing buildings had done so on non-residential buildings.

Wei et al. (2020) compared the IEQ category relative importance proportions developed between their two methods: (i) review of the 55 certification schemes and (ii) review of the nine peer reviewed studies. Wei et al. concluded that IAQ and thermal environment indicators were assigned slightly higher importance on average than acoustic and visual environment indicators, considering both the certification schemes and the rating systems developed in peer reviewed studies. However, the relative importance among the four categories of indicators was still nearly equal (most near 25%) using either method. Wei et al. stated this equal weighting of IEQ indicator categories occurred because there is a lack of scientific justification for assigning one IEQ indicator category higher importance than another. The authors therefore called for more research on the relative importance of the four considered IEQ indicator categories.

Possibly the most comprehensive attempt found in literature at creating a standardized framework to evaluate energy efficiency upgrades in residential buildings was performed by Basu et al. (2019). They conducted a comprehensive review and derived 34 "Energy Performance Indicators," (EnPIs) to help plan energy efficiency retrofit projects in multifamily homes in India. These indicators were split into six categories of factors influencing energy performance: (i) climate (based on geographic location); (ii) building envelope (characteristics of building surface areas); (iii) building services and energy systems; (iv) building operation and maintenance; (v) occupant-centric (occupants' activities and behavior); (vi) and indoor environmental quality. Self-administered ranking questionnaires were developed and distributed globally to individuals who were identified as either academic or industry practitioner experts in the built environment. Most of the expertise held by the consulted practitioners was in architecture and design. Experts in energy analysis, multiple fields of engineering, and energy efficiency/climate change policy work were also included. The questionnaires asked these experts to assign weights to the Energy Performance Indicators, prioritizing the indicators and indicator categories that are most important. Notably, consensus was not achieved, a finding reached independently by Wei et al. (2020) in their review and evaluation of a large number of related studies. However, Basu et al. (2019) did find that occupant-centric indicators were deemed most important by many of the experts with whom they engaged. This finding suggests how critical resident behaviors are likely to be when creating a framework for impact evaluation of building energy efficiency programs and projects. Taken collectively, our assessment of the recent literature informed our emphasis on empirical measurements of indoor conditions and in-home behaviors. This is in contrast to typical practice of modeling IEQ and occupant behavior based solely on building structural characteristics and static features of energy management systems. The current study contributes specifically to the selection of IEQ indicators, in-home energy use, and occupant energy-related behaviors in future frameworks for impact evaluation of building energy efficiency programs and projects. This study also contributes to guidelines for application of empirical measurements of IEQ, in-home energy use, and occupant energy-related behaviors in these frameworks.

2.2. Quality Criteria for Evaluating Metrics

Criteria have been developed and employed across a wide range of applications in diverse scientific fields to standardize evaluation of data quality and utility. Quality criteria may be applied to both qualitative and quantitative data. Three criteria that we considered relevant for our application (i.e., the assessment of metrics/indicators to be included in impact evaluations of home energy efficiency upgrade programs and projects) were selected or modified from criteria defined in health-focused and impact evaluation literature: (1) responsiveness, (2) representativeness, and (3) construct validity. The current study proposed definitions of these three criteria as they would be applied to metrics in energy efficiency upgrade frameworks (Table 2).

In epidemiological studies, *responsiveness* has been primarily treated as a method to evaluate measures used in questionnaire instruments. R*esponsiveness* was described in Terwee et al. (2007) as "a measure of longitudinal validity," and Windle et al. (2011) defined it as the ability (of a questionnaire) to detect changes over time. In the current study, *responsiveness* was adapted to mean how well an indicator captures changes in a home that are hypothesized to be linked to an initial change in energy efficiency in the home. The definition of *representativeness* used in this study derives from two concepts: (i) content validity – a criterion common in many studies evaluating environmental exposure metrics, among others (Dellinger & Leech, 2007; Hayashi et al., 2019; Seifert, 1995; Windle et al., 2011) – and (ii) reproducibility, as used by Windle et al. (2011) and Terwee et. al (2007). *Representativeness*, as defined in the current study, may be used to quantify how well an indicator comprehensively represents the domain that is being considered. In the application of assessing IEQ impacts of energy efficiency upgrades, *representativeness* can be applied to the spatial and temporal domains – i.e., how well does a short duration sample within a home represent a long-term period of interest (the IEQ conditions in the home on an average day), and how well does a sample from one location (one room) represent the larger region of interest (the entire home)? *Representativeness*, and how it is applied to the spatial and temporal domains, is explained in further detail later in this paper. The third proposed quality criterion, *construct validity*, has been commonly used in the context of epidemiology questionnaire instruments to assess how well an instrument "measure[s] the construct [it is] designed to measure," (Dellinger & Leech, 2007). Applying the concept of *construct validity* in the context of IEQ and home energy efficiency, this criterion may be treated as a means to describe how closely a given metric relates to human health. For example, if a study focused on occupant health measures indoor noise levels accurately and precisely, the noise data themselves may still have low construct validity if noise levels in homes do not have adverse effects on occupant health.

Table 2: Quality criteria for evaluation of IEQ and energy use metrics

Quality Criterion	**Definition**
Responsiveness	Extent to which a significant overall response is observed between metric before and after upgrade
Representativeness and Specificity	Extent to which data recorded for a metric during a short period, or from a sampling location encompassed within a region of interest, wholly characterize long-term conditions in the region of interest
Construct Validity	Extent to which metric corresponds to health impacts of upgrade on occupants

The current study did not evaluate *responsiveness* of metrics with respect to energy efficiency upgrades because most of the enrolled homes did not install upgrades during the study period. A comprehensive review of the *construct validity* (as defined in Table 2) of metrics that may be (and have been) used in energy efficiency frameworks is outside of the scope of this study. However, *construct validity* has been explored in the epidemiologic literature for airborne pollutants. High *construct validity* has justified measurement of the pollutants we selected for measurement in the present study. Wilson et al. (2016) performed a comprehensive overview of associations observed between IAQ indicators (including the ones we selected for this study) and occupant health. Wei et al. (2020) covered the frequency with which many other IEQ indicators (such as noise, light, and temperature) are used in studies to evaluate the effect of comfort-related parameters on human health, although they mention the need for more robust evidence regarding links between these IEQ indicators and health.

In this study, we focused on *representativeness* and *specificity*. These concepts, when applied to IEQ-related metrics, can provide insight on the following questions: (1) how representative are short duration IEQ measurements of long-term IEQ daily trends? (2) how representative are single-zone IEQ measurements of multi-zone IEQ? (3) what information is gained, if any, from longer-term or more spatially-resolved IEQ sampling? Practically, this work sought to shed light on questions like "When considering resource constraints on sampling period lengths, how long is long enough to adequately represent the general conditions within a home?" and "When considering resource constraints on availability of equipment and instruments, how well does measurement in a single location characterize the IEQ conditions within a household relative to multiple locations?"

3. Objectives and Hypotheses

Our study was guided by the following three objectives: (1) characterize temporal and spatial patterning of energy and IEQ metrics within and between homes, (2) assess correspondence between energy and IEQ metrics by temporal scale and spatial scale, and (3) use defined quality criteria to identify "fit-for-purpose" metrics for evaluating upgrades. Out of the three quality criteria defined in this study – i.e., *responsiveness, representativeness, and construct validity* – we focused on the evaluation of representativeness of metrics.

Analysis of indoor environmental quality indicators (IEQ) was stratified by season in this study. For each home, days were defined as cooling days (i.e., days when air conditioning was used), shoulder days (i.e., days when no air conditioning was used and minimal space heating was used), and heating days (i.e., days when moderate to high space heating was used). The methodology for defining seasons is explained further in Section 4.3.6. We hypothesized that: (i) occupants may be more likely to open windows on shoulder days; (ii) occupants would be more likely to keep their windows closed on heating and cooling days; (iii) occupants may engage in more indoor activities during heating days due to low outdoor temperatures in Colorado, and (iv) occupants may engage in less indoor cooking on cooling days to avoid the addition of heat to their home. As occupant behaviors can impact indoor environmental quality indicators, we hypothesized that certain seasons may require larger sample sizes to achieve representative samples of indoor environmental quality indicators.

4. Methods

4.1. Study Design

4.1.1. Study Area

The current study took place in Fort Collins, Colorado, a mid-sized city of approximately 160,000 people covering approximately 120 km^2. Fort Collins is situated directly east of the Colorado "Front Range," which is a mountain range within the greater Southern Rocky Mountains of North American that runs south to north in central Colorado. In 2020, the median reported household income was $70,474. In 2019, there were approximately 71,000 housing units, 55% of which were owner-occupied, and 45% of which were renter-occupied (United States Census Bureau, 2019). The area has a semi-arid climate characterized by mild winters and dry, hot summers that receive low amounts of rain, although afternoon thunderstorms can be common in late summer.

4.1.2. Epic Homes Program

The City of Fort Collins' Epic Homes program for residential energy efficiency is a suite of services built on the foundation of the Efficiency Works Homes program. Efficiency Works, formed in 2014, is a collaborative program serving utility customers of Platte River Power Authority's four co-owner cities in the northern Colorado Front Range area. Efficiency Works aims to increase energy savings by making energy efficiency offerings and rebates available to residential and commercial utility customers. Each utility designates funding to go toward the collaborative program, although funding is not transferred between communities; the funding provided by one utility is only used in the community that is served by that utility. Efficiency Works Homes is a branch of Efficiency Works that focuses on increasing the energy efficiency of homes; this includes single-family attached and detached, renter-occupied, and owner-occupied homes. The Epic Homes program is Fort Collins Utilities' specific residential energy efficiency program that offers additional services to Fort Collins Utilities customers on top of the standard assessment and rebate offerings of the Efficiency Works Homes program.

The purpose of the Epic Homes Program is to improve the energy efficiency of Fort Collins homes (while also improving the health and well-being of residents) by offering technical and financial assistance for home energy efficiency upgrades. Technical assistance is offered in the form of energy assessments. During an assessment a city-affiliated energy advisor evaluates several aspects of the home and suggests improvements that could be made to the home to allow for energy savings. These advisors then refer residents to possible rebates and low-interest on-bill loans that can increase the up-front affordability of upgrades. A "streamlined assessment" option is also available, through which advisors additionally connect customers with pre-approved in-network contractors who offer a pre-negotiated package price for efficiency upgrades. This connection with in-network contractors removes the need for the customer to acquire multiple contractor bids. A goal of the emergent Epic Homes program is to expand the served population from middle- to upper- income households (that tend to be owner-occupied) to additionally include low- to moderate- income households (that tend to be renter-occupied). Between June 2019 and August 2021, 1,348 homes were upgraded through the program, including 71 rentals (City of Fort Collins, 2021).

4.1.3. Recruitment and Study Duration

Owner-occupier and renter-occupier participants were recruited mainly via email. Participants and their respective homes were accepted into the study on a rolling basis once it was determined that the home met eligibility criteria. Eligible homes had access to wireless internet (allowing data uploading from IEQ and energy use monitors), were heated by natural gas, were served by Fort Collins Utilities, and had received energy assessments through the Epic Homes Program before or soon after entering the study. Houses heated by natural gas were chosen, as they were expected to be representative of a majority of the Fort Collins housing stock; an estimated 75% of homes in Colorado used natural gas for space heating in 2009, and 63% of homes in the Mountain North region of the United States (which includes Colorado) used natural gas for space heating in 2015 (U.S. Energy Information Administration, 2021). The current study aimed to gather environmental quality and energy use data within each recruited home for six to twelve continuous months. The protocols used in this study were approved by the Colorado State University Institutional Review Board (Protocol 19-9338H).

4.2. Data Collection

4.2.1. Indoor Environmental Quality Metrics

Indoor Environmental Quality (IEQ) metrics (with abbreviations and units shown in parentheses) included the following: fine particulate matter ($PM_{2.5}$; $\mu g/m^3$), a measure of total volatile organic compounds (TVOC; parts per billion (ppb)), and carbon dioxide (CO_2; parts per million (ppm)) concentrations, along with temperature (degrees Celsius), relative humidity (%), light intensity (lumens per square meter (lx)), and noise levels (decibels (dB)). These IEQ metrics were gathered continuously at five-minute resolution with commercial-grade indoor air quality (IAQ) sensing and monitoring devices (Omni, Awair, USA). The device determines $PM_{2.5}$ concentrations via a laser-based light scattering sensor (range: 0-1000 $\mu g/m^3$, accuracy: ±15% or ±15 $\mu g/m^3$), TVOC concentrations via a multi-pixel metal oxide gas sensor (range: 0-60,000 ppb, accuracy: ±10%), and CO_2 concentrations via a Non-Dispersive Infrared (NDIR) sensor (range: 400-5,000 ppm, accuracy: ±75 ppm or 10% of reading). Within each recruited home, three IAQ sensors were placed on walls in indoor living spaces: one in the kitchen, one in the living room, and one in a bedroom. Also, one IAQ sensor was placed on a wall in the garage, and one was placed outside (example in Figure 1). The outside sensor was typically placed within six feet of the home on a small post, with preference given to the backyard. All IAQ sensors were installed to be within typical breathing zones (three to six feet from the ground).

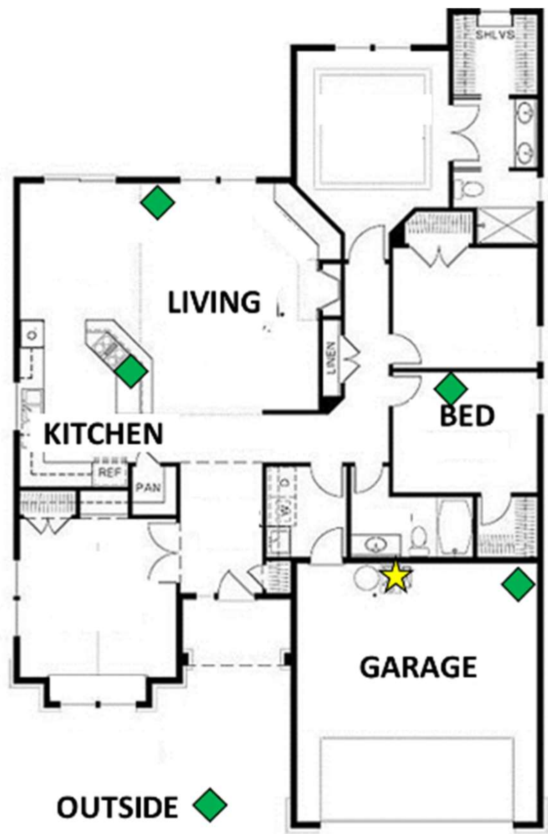


Figure 1: Example layout of home locations of commercial-grade indoor air quality sensors (green diamonds) and energy monitor (yellow star) indicated.

Continuous $PM_{2.5}$ measurements from the IAQ sensors were calibrated with concurrent, filter-based gravimetric measurements. Small ultrasonic personal aerosol samplers (UPAS) (Volckens et al., 2017) housed filters, which collected air samples over a three- to seven- day period upon each home calibration deployment. At each deployment, one UPAS was placed next to the living room IAQ sensor, one UPAS was placed next to the garage IAQ sensor, and one UPAS was placed below the outdoor IAQ sensor. The living room UPAS devices were placed within a 3-D printed, foam insulated case to mitigate any noise created during device sample collection. In addition to calibration prior to deployment, two in-field calibrations were conducted for all homes. The first in-field calibration occurred approximately seven to eight months after recruitment, and the second occurred nine to ten months after recruitment.

4.2.2. Energy Use

A residential energy monitor (Home Energy Monitor, Sense, USA) was used to monitor the electrical energy use of each home in real-time during the study period. Briefly, the home energy monitor is installed in the electrical panel of a home, allowing it to monitor electrical usage from all devices in a home at a one-minute resolution. The home energy monitor uses machine learning algorithms updated from the overall fleet of home energy monitors sold by the company and in use in U.S.-based homes (likely n > millions) to classify electrical devices in a home. The energy monitor subsequently assigns how much of the total energy consumption at any point in time is attributable to each classified electrical device. Once the home energy monitor identifies the power signature of an electrical device (typically, several days to weeks), it records the timestamps when the device is on, the power consumption at those times, and the overall energy consumption of the device. Natural gas usage data were also obtained from select, consenting homes to compare against energy monitor data used to measure heating and cooling behavior.

4.2.3. Energy Assessments and Questionnaires

Information on homes gathered by contractors during energy assessments included general home characteristics such as age/size of home, number of occupants in home, and limited information on appliances. Blower door tests were also conducted in each home, allowing for the calculation of CFM_{50}, or the airflow in cubic feet per minute (CFM) through a home when pressurized to 50 psi, and the calculation of ACH_{50}, the amount of air changes per hour (ACH) a home experiences when pressurized to

50 psi. CFM_{50} and ACH_{50} are standard measures of air tightness within a home. Questionnaires were administered to participants by researchers in the study to acquire additional information on home characteristics, such as when certain portions of the home were last upgraded. Acquired information from these questionnaires also included self-reported behaviors, such as window opening frequency and methods of heating and cooling. Questionnaires were also administered to determine whether any significant upgrades had occurred within homes, or if participants had vacated a home for extended periods of time.

4.3. Statistical Analysis

4.3.1. Data Management

The computer software, R (R Core Team, 2021), was used to handle all data gathered for this study and to create any of the included plots. All data were filtered to ensure no data outside of the range of deployment dates for each home (from periods of sensor testing) were included in the dataset used for analysis. IAQ sensors installed outdoors occasionally malfunctioned during periods of low temperatures. These malfunctions caused isolated temperature readings to exceed 10,000,000 ^{o}C during these cold time periods. Temperatures that exceeded 10,000,000 ^{o}C were thus omitted from the analyzed dataset.

4.3.2. Calibration of Real-Time $PM_{2.5}$ Measurements

Filter-based gravimetric measurements from all UPAS devices deployed within the study period were used to calculate a single gravimetric correction factor for the real-time $PM_{2.5}$ measurements recorded by the IAQ sensor. The use of pairwise correction factors (using a correction factor specific to each participant) was contrasted against the use of a single correction factor for all participants by Tryner et. al (2019). Tryner et al. concluded that a constant correction factor calculated from a random subset of participants performed better than using multiple pairwise correction factor values. As there were a small number of participants in our study, values from all participants were considered in the calculation of this correction factor. Two calibration deployments occurred for each home, but filter samples for which the UPAS collected less than 24 hours of data were not used for the calculation of correction factor values.

Two methods were tested to calculate a correction factor. The first calculated a pairwise correction factor for each sensor with the following equation:

Equation 1

$$\text{Correction Factor} = \frac{\text{TWA}}{\text{C}_{\text{filter}}}$$

Where TWA is the time-weighted average of all $PM_{2.5}$ measurements recorded by an IAQ sensor over the sampling period for which a filter sample was collected, and C_{filter} is the average concentration of $PM_{2.5}$ determined using gravimetric analysis of the filter with which the IAQ sensor was paired. The filter sample gathered in the living room was paired with the time-weighted average $PM_{2.5}$ concentration for each of the three indoor sensors in each home. We adopted this approach to sensor calibration because this study focused on relative trends and variability over time (and space) more than the absolute values of the IEQ metrics. Correction factor values calculated from replicate filter measurements (two calibration visits with valid data in some homes) were pooled with the other pairwise correction factor values for this and the following method.

The second method for calculation of a correction factor was running a simple model with TWA as the dependent variable and C_{filter} as the independent variable (both TWA and C_{filter} defined previously). The intercept of the simple model was set at zero. The resulting slope of the model would be used as the correction factor for all real-time measurements. The two methods of correction factor calculation were compared for performance before one method was chosen and used to calibrate IAQ sensor measurements prior to analysis.

4.3.3. Autocorrelation

Time series data are subject to autocorrelation, which means measured values are influenced by values that occurred at previous points in time. Significant autocorrelation is common with real-time air quality datasets because air pollutants that are released within a region or indoor space at a given sampling time may not have dissipated before the subsequent sampling time. When this occurs, subsequent data points cannot be considered independent. Autocorrelation is an issue with model creation, as autocorrelated data can result in misleadingly high sample sizes. Autocorrelation can be mitigated in a dataset by accounting for a calculated autocorrelation coefficient within model algorithms. It can also be mitigated by averaging data over a sufficiently long time period to result in independent, averaged data points before using them to develop a model. The required averaging period can be determined by measuring autocorrelation using an autocorrelation factor (ACF), at multiple time "lags." This is done by "lagging" a dataset by one time increment (translating every data point to have a timestamp of $t+1$), then calculating the Pearson correlation of the lagged dataset with the original dataset. The resulting calculated correlation value is the autocorrelation factor at a time lag of 1. This process is repeated for multiple lags (lagging by two time increments, then three, then four, etc.).

Autocorrelation factors can range from -1 to 1; -1 implies a strong negative autocorrelation (increasing time series values are associated with decreasing subsequent time series values); +1 implies a strong positive autocorrelation (increasing time series values are associated with increasing subsequent time series values), and 0 implies no autocorrelation (time series values have no directional association with subsequent time series values). Autocorrelation can be visualized with correlogram plots, in which autocorrelation factors are plotted on the y-axis vs time lag values on the x-axis (examples from the current study in Figure 2). A two-sided confidence interval for the autocorrelation factor is created based on the amount of tested data points for the correlation calculation at each time lag. Autocorrelation within a dataset is considered insignificant once the autocorrelation factor is contained within the bounds of the confidence intervals (Brockwell & Davis, 2016; Gerbing, 2016; SAGE, 2017). The difference between high and low confidence bounds in a correlogram becomes greater at higher time lag values. This increasing difference occurs because the number of pairs used to calculate the correlation between an original dataset and its lagged counterpart decreases by one as the number of time lags between the original dataset and its lagged counterpart increases by one. This in turn causes an increase in the magnitude of the standard error value used to calculate confidence. The detailed methodology used in this study for confidence interval calculation can be observed in a lecture published at the University of North Carolina (Weiss, 2012).

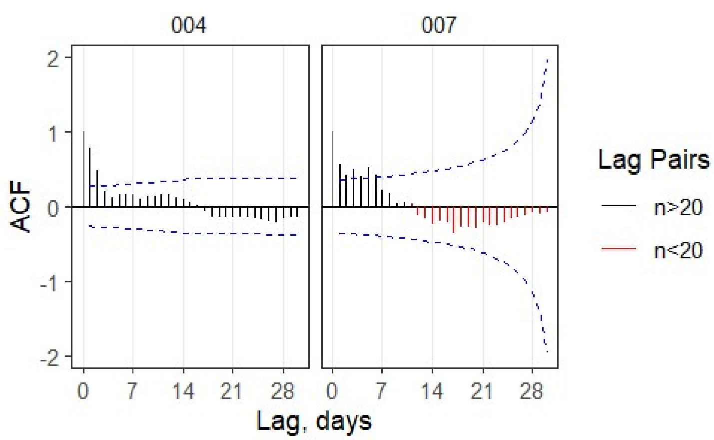


Figure 2: Example correlograms for TVOC data recorded in Home 4 (left) and Home 7 (right) in the current study. Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

No models were created in the current study. However, to create boxplots of independent data points for IEQ indicator values, the averaging period length that resulted in independent data points needed to be determined beforehand. This averaging period was expected to be different between IEQ metrics. Autocorrelation analysis was thus performed for each IEQ indicator at time lags ranging from one day to thirty days. This was performed separately for datasets collected over each season from each sensor within each home. The number of time lags required for the autocorrelation factor to reach insignificance was considered in the definition of the required averaging period for each IEQ indicator. Calculated autocorrelation values were also compared between air pollutants to understand the difference in the pollutant behaviors post-emission. A pollutant experiencing higher autocorrelation factor values (closer to +1) for many time lags implies the pollutant takes longer to dissipate once emitted than a pollutant for which the autocorrelation factor value quickly drops from +1 to 0 over a small number of time lags (Luoma & Batterman, 2000). Significant negative autocorrelation values were not expected on a daily time-lag scale, as increasing concentrations of the air pollutants measured in this study on one day would not be expected to be associated with decreasing concentrations on subsequent days.

A dataset (from one sensor in one season) was omitted from autocorrelation analysis in the current study if it the time period for the dataset was not at least twenty-five days in length, as correlations should only be calculated between datasets that contain at least twenty data points (StatSoft, 2011). Ideally, autocorrelation is only calculated from samples that are not missing values; however, datasets without any missingness were uncommon in the current study (as is the case with most long-term air quality studies). A dataset was omitted from autocorrelation analysis in the current study if it was missing more than 11.11% (1/9) of the days within its corresponding time period, per the omission criteria used by Zhao et al. (2018) in their development of a model reliant on calculated autocorrelation values. Lastly, stationary data (data without long-term directional trends) is required to allow for proper autocorrelation analysis (SAGE, 2017). The number of time lags before autocorrelation factor values reach insignificance can be impacted if the analyzed data is non-stationary. If, for instance, an IEQ indicator value experiences a constant increase or constant decrease over the season during which autocorrelation analysis is performed, autocorrelation factor values will likely stay high over the tested number of lag days. High autocorrelation factor values occurring in the case of non-stationary data are due to long term changes in the environment (such as ambient air pollutant levels slowly decreasing due to a change in season), not the failure of an emitted pollutant to disperse due to strong physical persistence or low air change rates. Most modeling efforts aim to accurately and precisely characterize pollutant emission and dispersion patterns. Modeling studies will sometimes "de-trend" data by subtracting a smoothing function from all days before determining autocorrelation factor values. Alternatively, some modeling studies de-trend data by calculating autocorrelation factor values of the difference between values recorded on subsequent days, instead of finding autocorrelation factor values of the raw values. In the current study, autocorrelation analysis was used to determine the averaging periods required for independent samples, not to predict future values. Therefore, autocorrelation analysis of transformed (de-trended) IEQ indicator values were not as pertinent as autocorrelation analysis of raw IEQ indicator values. Instead, datasets for which autocorrelation remained significant after thirty sampling days were excluded from consideration for determining averaging periods. Overall, considering the three autocorrelation omission criteria discussed above, 24.6% of $PM_{2.5}$ datasets, 26.2% of TVOC datasets, and 32.3% of CO_2 datasets were excluded from consideration for determining averaging periods. Omissions of a dataset from autocorrelation analysis did not imply omission of the dataset from other analyses performed in this study.

4.3.4. Summary Statistics

Time-weighted averages (TWAs) of IEQ indicator values were calculated from real-time data collected over the entire monitoring period for each indoor sensor in each home. These TWAs were used for comparison to other studies. Coefficient of variation (COV) values were also calculated for each IEQ indicator to compare the variation between indicators. COV (calculation shown in Equation 2) can be used to compare variation in metrics around their respective means when the metrics have different units, ranges, and mean values.

Equation 2

$$\text{COV} = \frac{standard\ deviation}{mean}$$

A higher COV implies more dispersion around the mean, proportional to the magnitude of the mean.

We compared IEQ indicators seasonally between room, first time-averaging each data series using the averaging periods determined with autocorrelation analysis.

4.3.5. Time Series and Diel Analysis

We visualized data using diel plots showing hour-of-day averages. Diel analysis was performed because IEQ metrics have been shown to exhibit patterns that repeat daily (Maciejewska & Szczurek, 2015). The 95% confidence intervals were used to represent variation in the data. We used Spearman correlations to evaluate if correlations existed and/or varied by time of day between the same IEQ metric measured in more than one room. We evaluated the Spearman correlations between 1-h, time-averaged values matched in time across room-pairs (e.g., living room and kitchen), resulting in 24, between-room correlations for each home. Spearman correlations are a measure of the strength of the ordinal relationship between two variables. A Spearman correlation value near positive unity (+1) implies that as one variable increases, the other does as well (and as one variable decreases, so will the other). A Spearman correlation value near negative unity (-1) implies that as one variable increases, the other decreases, and vice versa. This methodology was also used to explore diel correlations between IEQ metrics recorded within each room. Autocorrelation was present in the data. Spearman correlations were performed for the relationship between rooms and separate metrics, not consecutive data points within a given location. Therefore, no adjustments for autocorrelation were deemed necessary to consider the resulting Spearman correlation values (Afshar-Mohajer et al., 2020).

4.3.6. Defining Seasons

IEQ indicators within homes can be significantly impacted by human behaviors. We integrated measures of occupant heating and cooling behavior, as determined by the home energy monitor, to classify days based on space conditioning behaviors. Past studies that developed air quality models have accounted for space heating and cooling behaviors through the proxy measures of *heating degree days* and *cooling degree days* (Fazli & Stephens, 2018). These measures consider the daily difference between indoor and outdoor temperature, assuming that the amount of energy used to heat or cool the home will be linearly related to the magnitude of this temperature difference. Our study proposed a more direct measure of space heating, using high-resolution energy consumption data recorded by the home energy monitors for heating and cooling devices. Days were classified into three categories: heating, cooling, and shoulder. This approach of classifying days based on space heating behavior is similar to the approach used by Deng et al. (2021). In their study of indoor environmental quality in 85 schools across the United States, Deng et al. classified days as heating days or non-heating days based on the daily operation of the HVAC system. In our study, days were classified based on the daily fraction-on time of the home's primary heating and cooling devices. For instance, if on a given day, a heating device was detected in-use for 6 hours, the daily heating fraction-on time was 0.25. The home energy monitor data were used to identify heating devices (e.g., furnace) and cooling devices (e.g., air conditioning, or AC). Heating devices were present in all homes, while cooling devices where present in only some homes. For homes in which the home energy monitor identified an air conditioner, the participant confirmed they owned either central AC or a window AC unit. All homes in this study were heated by natural gas. We assumed that the home energy monitor detects a heating device because there are electrical elements that activate when a heating device powered by natural gas is operating. It was also assumed that the energy output of a heating device in operation is relatively consistent. This consistency assumption implies that if a device is operated for one hour on a given day and two hours on a subsequent day, twice the amount of heating energy is supplied to the home on the subsequent day. Therefore, the daily fraction-on time of heating devices identified by the home energy monitor was used as a proxy for indoor space heating.

It was assumed that the amount of heating-related natural gas consumed in each month would be proportional to the amount of time the primary heating device operated within the home in that month. Therefore, heating-related natural gas consumption was used to determine a threshold of daily heating device fraction-on time. We would classify each day with fraction-on time above this threshold as a heating day. To establish this daily heating threshold, we evaluated average monthly natural gas consumption for space heating by Colorado residents (Figure 3), using data from the U.S. Energy Information Administration (EIA) (2021). Monthly residential natural gas consumption data from the EIA were gathered for the state of Colorado from January 2016 to December 2020. From this analysis, residential natural gas consumption for heating was greatest in December. Natural gas consumption during traditional summer months (June, July and August) was <15% of the average December consumption value. Thus, 15% was chosen as a threshold for defining the fraction-on of heating devices for a heating day.

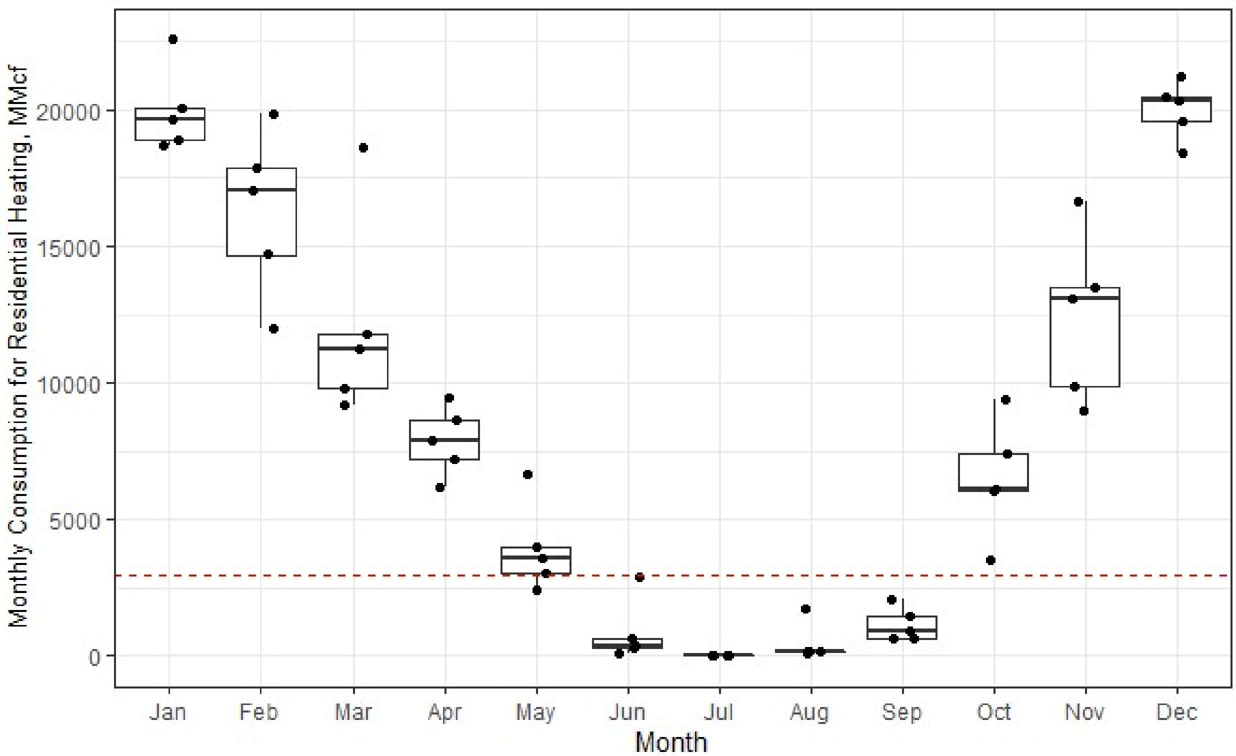


Figure 3: Estimated monthly natural gas consumption by Colorado residents for heating, from 5 years (January 2016 to December 2020) of U.S. Energy Information Administration (2021) *data. Dashed line shows chosen threshold of 15% of average consumption for month of greatest 5-year average consumption (December). MMcf = Million cubic feet.*

For each home in the current study, the average daily fraction-on time was calculated for each month. The month with the maximum average daily fraction-on time was identified, and 15% of this fraction-on value was defined as the home-specific threshold for a heating day. For example, in a given home, if the average daily fraction-on time for the identified heating device was greatest in the month of December, equal to 0.5, the heating device was on for 12 hours on an average December day. Days for which the heating device's fraction-on time was greater than 0.075 (i.e., 15% of 0.5) were defined as heating days for that home. Days when the heating device fraction-on time was below the 15% threshold, and when air conditioning device usage was not detected (i.e., air conditioning device fraction-on time = 0) were defined as shoulder days. Days when the heating device fraction-on time was less than the 15% threshold, and when use of an air conditioning device was detected (i.e., air conditioning device fraction-on time > 0) were defined as cooling days. The purpose of using this threshold value was to avoid the constraint of categorizing days by month of year, instead categorizing days by resident behavior. Resident behavior may affect IEQ variables more significantly than weather or other outdoor variables commonly used in IEQ models. The heating device 15% threshold was chosen to approximate, on average, what amount of daily heating may be used on days when behaviors are likely to be different – e.g., when windows are open more commonly throughout the day. The air conditioning device threshold was set at a fraction-on time of zero. This was because, for each home, the daily fraction-on time of the primary air conditioning device was observed to vary less throughout the year than the primary heating device fraction-on time. We hypothesized that: (i) occupants may be more likely to open windows on shoulder days; (ii) occupants would be more likely to keep their windows closed on heating and cooling days; (iii) occupants may engage in more indoor activities during heating days due to low outdoor temperatures in Colorado, and (iv) occupants may engage in less indoor cooking on cooling days to avoid the addition of heat to their home. For all homes, each individual day was classified as either heating, cooling, or shoulder when energy data was available.

For many homes, more than one device of each type (heating and air conditioning) was detected. This was likely due to the identification of secondary devices – e.g., small space heaters – or due to misidentification. When more than one heating or air conditioning device was detected within a home, the heating or air conditioning device that experienced the highest daily fraction-on was identified as the primary heating or cooling device. A few exceptions occurred when two devices had similar fraction-on values. In these cases, we made what we deemed to be the more rational choice. For instance, if one heating device had higher fraction-on time in December and January while the other heating device had higher fraction-on time in July and August, the device with higher fraction-on time in December and January was classified as the primary heating device. The usage of these primary heating and cooling devices was then referenced when determining the season identification (cooling, shoulder, or heating) for each day.

Table 3: Methodology for defining the season to which each day belongs for each home using heating and cooling behaviors with expected behaviors that may affect IEQ variables.

Season	Definition	Expected Behaviors
Heating	Heating device fraction-on time > 15% of average daily fraction-on value in month with maximum average daily fraction-on value	• Windows frequently closed • Possibly elevated indoor resident activity
Shoulder	Heating device fraction-on < 15% of average daily fraction-on value in month with maximum average daily fraction-on value and no air conditioning detected	• Windows frequently open
Cooling	Heating device fraction-on < 15% of average daily fraction-on value in month with maximum average daily fraction-on value and air conditioning detected	• Windows closed • Possibly reduced cooking activity (fewer hot meals)

Homes appeared to follow three patterns of heating and cooling behavior based on our analysis; the homes were therefore categorized into three groups (Figure 4). Group 1 homes were defined by frequent changes between season, switching constantly between cooling, shoulder, and heating days over the course of the monitoring period. These Group 1 homes lacked any definable continuous season periods. Group 2 homes lacked continuous cooling periods. Instead, Group 2 homes were typically characterized by continuous shoulder seasons during the initial and final days of their monitoring periods, and continuous heating periods during the middle potion of their monitoring periods. Most of the days within Group 2 monitoring periods were classified as heating days (Table 4). Finally, Group 3 homes were charactered by continuous cooling periods in the beginning portion of their monitoring period, followed by a less consistent shoulder period (interspersed with cooling and heating days), followed in turn by an extended continuous heating period. As with Group 2 homes, most of the days within Group 2 home monitoring periods were classified as heating days (Table 4).

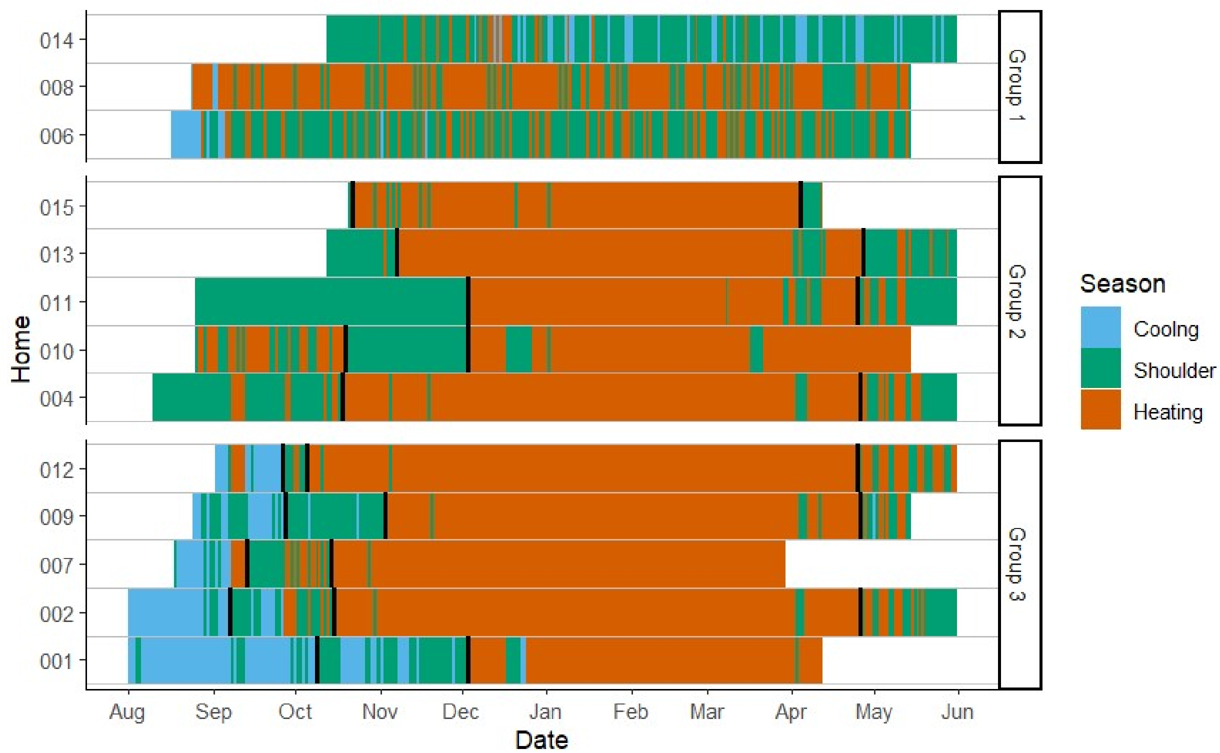


Figure 4: Categorization of days into heating and cooling behavioral-based seasons over entire monitoring period for each home. Colors display the category of day. White space symbolizes dates without data. Black lines define start and end dates of continuous periods used for representativeness analysis. Home "grouping" distinguishes homes with definable continuous cooling periods (Group 3), without definable cooling periods (Group 2), and without any definable periods (Group 1). Homes 3 and 16 were not included due to lack of home energy monitor data. Months span from August 2020 (Aug) to June 2021 (Jun).

Home energy monitor data for Homes 3 and 16 were not retrievable and were therefore not used for data analysis that was disaggregated by season. Days for which IEQ data was recorded, but no energy monitor data was available, were also not used for data analysis that was disaggregated by season. In the questionnaires, all homes except one (Home 4) reported, that they had, and used, air conditioning. Therefore, homes in Group 2 reported having air conditioning, but did not seem to use it during the monitoring period. In Homes 13 and 15, this may have been because their enrollment date occurred after traditional cooling periods. These two homes enrolled in October, when many Colorado residents are less likely to use air conditioning than in the summer months of June, July, and August. However, the lack of continuous cooling periods in Homes 10 and 11 suggests real differences in seasonal air conditioning behaviors between these two homes in Group 2 and homes in the other groups (i.e., Group 1 and Group 3).

Table 4: Percentages of days that were categorized into each season for each home group.

Home Group	Season		
	Cooling	Shoulder	Heating
1	7%	54%	39%
2	0%	31%	69%
3	13%	17%	70%

Group 1 homes exhibited heating and cooling energy use patterns that were challenging to interpret. For example, Home 14 had minimal heating use and continuous AC use throughout the monitoring period, even during winter months. Homes 6 and 8 exhibited consistent cycling between heating and shoulder days throughout the study period. This is unlikely to be representative of their actual heating behavior even if the participants desire to keep their home at a constant high temperature, as the amount of heating would be expected to increase in colder months if a constant temperature was desired. Also, the shoulder days defined for these homes were at unlikely periods for which residents would be opening windows (e.g., Homes 6 and 14 experienced multiple shoulder periods within the typically cold Colorado winter months of January and February). The method used in this study for determining energy use for heating and cooling appeared to be inadequate for these homes and would warrant further investigation in future work. Therefore, we excluded Group 1 homes from further analyses that involved disaggregating results by season.

When continuous seasons were defined for analysis, most Group 3 homes had shoulder seasons that were similar in length (ranged from 31 to 55 days in length for four of the five homes). Heating dates encapsulated the traditional winter period for Colorado (December 21st through March 21st), but for many homes, heating days occurred earlier (e.g., early October) and extended later (e.g., April). For homes in which the end of the heating season was observed, the start dates of the continuous heating season (median: October 22; range: October 5 to December 3) varied more between homes compared to the end dates (median: April 26; range: April 4 to April 27). This difference in variation suggests that seasonal transitions in resident space conditioning behavior may be more difficult to predict in the shoulder-to-heating transition than the heating-to-shoulder transition.

When time series data were disaggregated by season, heating days were grouped with other heating days, shoulder days were grouped with other shoulder days, etc. Data was also disaggregated by season for autocorrelation and representativeness analyses. However, autocorrelation and representativeness analyses required continuous data; subsequent data points used in analysis had to be collected from days that were collected subsequently to each other during sampling periods. Start and end dates were therefore identified for each season within each home for autocorrelation and representativeness analysis. If a home identified the use of an air conditioner in the beginning of the monitoring period, the initial days were identified as cooling days. The start of a shoulder season was identified for each home when the first eight consecutive shoulder days occurred. If a home did not have air conditioning, the beginning of its monitoring period was identified as a shoulder season. The start of a heating season was identified for each home when the first eight consecutive heating days occurred. The end of the heating season for a home (if the end of the heating season occurred within the monitoring period) was marked at the end of the final span of eight consecutive heating days. If the heating season did end prior to the final date of a monitoring period within a home, days following the end of the heating season were not used in representativeness or autocorrelation analyses; time periods that followed heating seasons were often too short to analyze.

4.3.7. Testing Normality

Assumptions of normality were required for evaluation of spatial representativeness and specificity in this study. As such, QQ plots were generated and visually inspected to evaluate how well a given distribution of data fit the assumption of normality prior to performing calculations. Data distributions were evaluated against log-normal, gamma, and Weibull distributions to determine the best fit for transformation.

4.3.8. Representativeness and Specificity

4.3.8.1. Literature Review on Representativeness

We define spatial representativeness of IEQ metrics as the extent to which measurements from a single sampling location capture trends and magnitudes that characterize average conditions within the greater spatial extent of interest. As defined here, spatial representativeness is not well explored in IEQ literature. Between-home or between-building variability is often considered in IEQ studies, but between-room variability within homes is less studied, often due to a lack of measurements in multiple rooms. When between-room variability is explored, it is often done at low temporal resolution; only the average values of an IEQ metric gathered over an entire monitoring period are compared between rooms. If high temporal resolution (e.g., hourly) data is gathered, it is only gathered for a short period of time, given resource and time constraints. Comparison of data at high temporal resolution is important to evaluate if trends in data – not just the overall means – from different locations are fully characterizing the region of interest. Diel trends for IEQ data in homes, for instance, may help identify potential sources of poor IEQ within a home, especially if combined with energy monitoring data that indirectly reflects indoor activities and behaviors of residents (e.g., cleaning or cooking).

There has been recent growth in studies that gather high resolution data from multiple locations within buildings and consider the variability between datasets. In a study on the variability of office building IEQ, Szigeti et al. (2017) looked at the between-office spatial variability of high-resolution particulate data in multiple buildings across Europe using Intra-Class Correlation Coefficients. Szigeti at al. compared spatial variability to temporal variability as well. Beko et al. (2016) measured interzonal air exchange rates within one home in real-time to evaluate how pollutants would be expected to travel between rooms under different scenarios. This study (Beko et al., 2016) specifically looked at impacts of different "source rooms" on other locations within the home under multiple conditions (e.g., source room door open/closed or home unoccupied). Wan et al. (2011) measured real-time living room and kitchen particulate matter levels in 12 homes during gas-stove-powered heating events. The authors measured three size fractions of particulate matter: ultrafine particles (14.6 to 100 nm in diameter), accumulation mode particles (100 to 661.2 nm in diameter), and $PM_{2.5}$. Wan et al. measured the concentrations of these three particulate matter size fractions in both rooms during, and directly following, gas-stove-powered heating events. In another study on cooking-related $PM_{2.5}$, Xiang et al. (2021) measured real-time $PM_{2.5}$ in the bedroom, living room, and kitchen in an apartment in Seattle, Washington. Real-time $PM_{2.5}$ measurements were recorded during cooking events under several different controlled ventilation scenarios. A recent study in Italy measured temperature, relative humidity, $PM_{2.5}$, and VOCs at high resolution in the kitchens and bedrooms of two homes. Data was recorded over one, two-week period in each of three seasons (winter, spring, summer) during the COVID-19 lockdown period (Pietrogrande et al., 2021). The exact resolution of measurements in Pietrogrande et al.'s study was not reported, but results showed that the authors had collected data of at least hourly resolution. Most of the studies we found that included between-room comparisons of measured IEQ indicators involved only means and standard deviations. Pietrogrande et al.'s (2021) study was the only study we found that evaluated correspondence between rooms on a high (hourly) resolution basis. Their study compared hourly concentrations of the kitchen and living room for one sampling day.

Even in studies where data have been gathered at high resolution (i.e., hourly for temporal resolution; multiple rooms for spatial resolution) in homes, these data have not been used to evaluate or demonstrate potential applications and value that higher resolution data may offer. For instance, would a practitioner be able to identify poor IEQ conditions within a home if only sampling from one room? Which room would be best to sample from, if only one sensor were available per home? What information is gained if only one room can be sampled for only a few days?

Studies have been performed to determine the optimal locations of monitoring stations within an outdoor air quality monitoring network (Caselton & Zidek, 1984; Perez-Abreu & Rodriguez, 1996; Silva & Quiroz, 2003). These studies used a measure of information called relative entropy (explained in detail in the following sections) to evaluate what they called the "effectiveness" of a network. The goal of these studies was to determine the arrangement of monitoring stations that would result in the most information gathered using the least stations. Osses et al. (2013) expanded on this concept, renaming the "effectiveness" index to a "specificity" index and using the index as a measure of how difficult it would be to gather data from a network without including a given sensor. This measure essentially determines how irreplaceable a given sensor is for the network. The authors then defined a "representativity" index, which they asserted was a required complement to the specificity index. The representativity index measured how well a sensor characterized general conditions within the region of interest by considering the change in uncertainty of a network when adding the given sensor to the network. These measures of specificity and representativity, referred to as "spatial specificity" and "spatial representativeness" respectively in our study, were slightly modified and used to evaluate the spatial resolution of the data gathered within enrolled homes.

We define temporal representativeness of IEQ metrics as the extent to which a sampling period captures trends and magnitudes that characterize average conditions in the long-term period of interest. As defined here, temporal representativeness has also not been well explored in IEQ literature. However, some measures have been introduced to evaluate the representativeness of samples. In their study of personal exposures to nitrogen dioxide and sulfur dioxide, K. Lee et al. (2004) perform paired t-tests, comparing the mean and 95th percentile of short term (one- to fourteen- day) periods to the respective long-term (fourteen-day) periods from which they were extracted. The authors used the results to determine if there was a significant difference between the summary statistics of the sample and the long-term period. This approach is uncommon in the literature, likely because paired t-tests require two independent samples for comparison, an assumption that may have been compromised in the aforementioned example.

Luoma & Batterman (2000) included representativeness as one of the five indicators of data quality, and write about how variability negatively impacts representativeness of sample measurements. The authors modeled how ventilation of a room can impact autocorrelation of IEQ indicators, and thus the variability of the indicators. Luoma & Batterman established relative standard deviation, or RSD (the standard deviation of the mean divided by the standard deviation of sample measurements) as a measure of variability. The authors applied this RSD measure to real-time seven-hour samples of particulates, bio-aerosols, and CO_2 recorded during each day of a one-week period. E.G. Lee et al. (2008) expanded off of Luoma & Batterman's study and used RSD (using the name of standard deviation ratio, or SDR) to estimate the amount of time required to achieve a representative sample of a tracer gas. E.G. Lee et al. use a targeted precision value as a threshold to define a representative sample. However, using the measure of SDR/RSD for analysis requires stationary data and samples that are assumed to have the same mean as the monitoring period during which the samples were recorded. As such, the SDR/RSD measure was only applied to short-term monitoring periods (a few hours at most) in the aforementioned studies. Also, neither this SDR/RSD method, nor the paired t-test method, can contrast the diel structure (hour-of-day averages) of sample data to the time-structure of the long-term period the sample is attempting to characterize. The SDR/RSD method and paired t-tests can only contrast means.

Relative entropy is the measure of information used to derive the metrics of spatial representativeness and specificity in the studies mentioned previously. Few studies have used relative entropy to measure temporal representativeness. Outside of the field of IEQ, Stanley et al. (2018) used relative entropy to determine how long GPS data should be gathered from the mobile phone of participants before the data is considered representative of typical mobility patterns. The threshold for representativeness was defined when information was no longer gained from the addition of another day of data. A follow-up study (Yoo, 2019) was then performed including participants from a broader span of socioeconomic and demographic backgrounds. Yoo aimed to determine if representativeness of different sample lengths varied by individual. Only one study, performed by Maciejewska & Szczurek (2015), was found applying relative entropy as a measure of temporal representativeness for IEQ samples. The authors recorded real-time measurements of temperature, relative humidity, and carbon dioxide in a lecture theatre and computer laboratory over several months. Maciejewska & Szczurek then calculated the representativeness (as derived from relative entropy) of a short sample to the month-long period from which it was extracted. The authors reported how representativeness varied as the length of the sample was increased from one day to the entire month. Maciejewska & Szczurek also defined a threshold for representativeness and determined how long a sampling period for each of these metrics needed to last to be considered representative of long-term periods ranging from one to nine months. Based on their defined threshold, the authors concluded that carbon dioxide measurements were required to be approximately 20% of the length of a long-term period of interest to be representative. Maciejewska & Szczurek reported that temperature and relative humidity measurements required longer sampling periods – roughly 30-50% of the length of the long-term period of interest – to be considered representative. Our study modified the methods used by Maciejewska & Szczurek (2015) to calculate the representativeness of short IEQ samples to the long-term period from which they were extracted within homes. The methods for this calculation are explained in more detail in the following sections.

4.3.8.2. Relative Entropy

Entropy is a measure of the uncertainty within a data distribution. The Kullback-Liebler Divergence, also called relative entropy, is a measure of the similarity between two probability distributions. This measure is often used in machine learning applications to determine the quality of models (Bousquet, 2008; Kowalski et al., 2014; Kullback, 1959). The below definition of relative entropy was considered for the measures used in this paper:

Equation 3

$$\text{Relative Entropy}, D(p_s||q) = \int p_s(x) \ln \frac{p_s(x)}{q(x)} dx$$

Or in the discrete form:

Equation 4

$$\text{Relative Entropy}, D(p_s||q) = \sum_{x \in A} p_s(x) \ln \frac{p_s(x)}{q(x)}$$

Where *$p_s(x)$* is the probability distribution of values from the considered sample, *$q(x)$* is the probability distribution of values from the corresponding long-term dataset, and *A* (in the discrete case) is the domain of possible values for *X*. Relative entropy is always positive; the larger the value of relative entropy, the more different the two probability distributions. Relative entropy is only zero if the two considered distributions are identical. In this study, measures derived from relative entropy were used to evaluate both spatial and temporal representativeness.

4.3.8.3. Temporal Representativeness

Temporal representativeness was measured using relative entropy in a similar fashion to Maciejewska & Szczurek (2015) by creating a probability distribution from sample data. The sample data's probability distribution was then compared to the probability distribution of the long-term monitoring period dataset from which the sample was extracted. The following equation was used to create an intuitive representativeness index that varied from 0 to 1:

Equation 5

$$\text{Temporal Representativeness}, \mathrm{R}_{\mathrm{T}}^{\mathrm{s}} = 1 - \frac{\mathrm{D}_{\mathrm{s}}}{\mathrm{D}_{\max}}$$

Where, for a given IEQ metric (e.g., $PM_{2.5}$, TVOC), D_s is the relative entropy of a sample *s*, and D_{max} is the largest relative entropy value (least representative) of all samples considering all sampling lengths, homes, rooms, and seasons.

For each sensor within a given home and room, the long-term monitoring period, against which the representativeness of a sample was measured, was defined as the entire monitoring period (generally six to ten months) of the given sensor. The health impacts associated with exposures to the indoor air pollutants measured in this study are typically outcomes of chronic and non-communicable diseases (e.g., cancer; cardiopulmonary illnesses); thus, the exposures applied in exposure-response models that include these pollutants are treated as chronic exposures. Even if the actual measure of exposure is much shorter (e.g., one to several days, multiple weeks, or even a year-long measure), the exposure is treated as a measure of chronic exposure (i.e., occurring at the measured level over the lifespan or over long periods of time) (International Energy Agency, 2017). Therefore, an ideal analysis might use an entire calendar year to encompass most conditions and behaviors that a home experiences. As such, we took our longest continuous period available in the dataset (six to ten months, depending on the home) to be the basis for evaluation.

Representativeness was calculated and disaggregated by season. It has been noted in literature how the changing of season can affect behaviors (Du et al., 2020), and corresponding changes in ventilation, cooking, and heating habits, can cause significant changes in emissions/infiltration of pollutants, and other measured metrics. Thus, we hypothesized that certain seasons may require larger sample sizes to achieve representative samples. For each home, the long-term monitoring period of six to ten months was disaggregated into the three considered seasons of heating, shoulder, and cooling (only heating and shoulder if air conditioning was not used in the home). Samples of lengths ranging from one day to 28 days were then extracted from each considered season within the long-term monitoring period. A sample length of one day was chosen as the minimum sample length, as this sample length would allow for the capture of diel trends. A sample length of 28 days was chosen as the maximum sample length. For typical in-home environmental quality evaluations, it was assumed infeasible for a practitioner to acquire hourly resolution air quality samples lasting longer than 28 days within a home.

For each sample length considered within the range of one to 28 days, samples were extracted via a moving window technique without overlap (exemplified in Figure 5). Partial samples that extended past the end of the season were ignored. For example, assume a given sensor had a heating season that was 65 days in length. When calculating how representative a ten-day sample recorded during the heating season is of the data collected over the sensor’s long-term monitoring period of ten months, six samples would be considered (65/10 = 6.5, rounded down to 6). A separate relative entropy value would be calculated for each of the six samples. A significant amount of data at the end of a given season is not considered for larger sample sizes when using this sampling window technique, as the ending partial sample windows are not included. In the aforementioned example, the final five days in the 65-day heating season would not be considered in any samples. To mitigate this problem, an *overlapping* moving window sampling technique was considered, where the start of each sampling window is shifted only one day later than the start of the previous sample. Using the overlapping sampling technique with the previously assumed heating season would have resulted in 55 samples (65-10 = 55) that spanned the entire season. However, concerns arose over the lack of independence between overlapping samples when considering this overlapping moving window technique. Therefore, the non-overlapping window sampling technique was used in the temporal representativeness analysis.

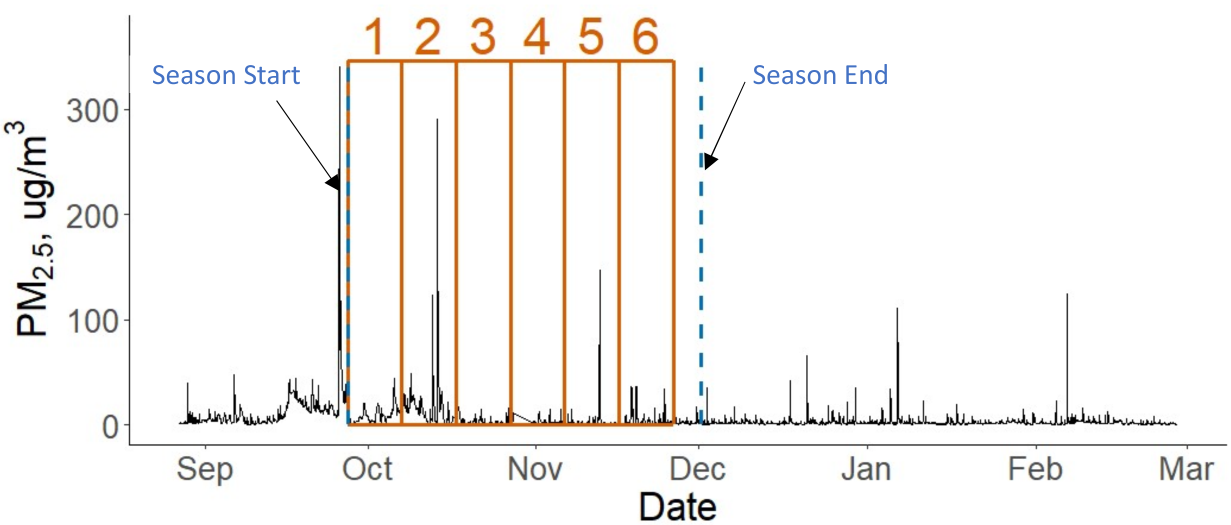


Figure 5: Example of sample extraction process for $PM_{2.5}$ data gathered from one sensor in a season. Numbered orange boxes represent extracted sample periods. Blue dashed lines represent the start and end of the season. The date range is the entire (long-term) monitoring period. This example shows six, 10-day samples extracted from a 65-day season within a six-month long-term period. Note that data period at the end of the season after the sixth sample is not used, as it is less than ten days in length.

Two methods of preparing the probability distribution functions of a sampling period and its respective long-monitoring period for comparison were described in Kowalski (2014). These two methods, as used in the current study, are visually exemplified in Figure 6. Using the histogram method – i.e., categorizing time series values into bins based on their magnitudes (not their time value) – would result in what we call a “magnitude-based” comparison of two distributions. Magnitude-based comparison would not consider the original temporal ordering of the data. Meanwhile, by splitting the time series into bins defined by the time value and averaging the magnitude of the time series value within each bin, one is able to compare the “time-structure” of two distributions (Bandt & Pompe, 2002; Kowalski et al., 2014). When conducting a time-structured comparison of two distributions, the absolute average magnitudes that are recorded within each time-defined bin are not directly compared, as the values within each distribution are normalized to achieve a probability distribution. However, relative values (such as the maximum average value divided by the minimum average value) for each distribution would remain unchanged when converting the distribution to a probability distribution. These relative values would therefore be considered when comparing the two distributions via a relative entropy calculation.

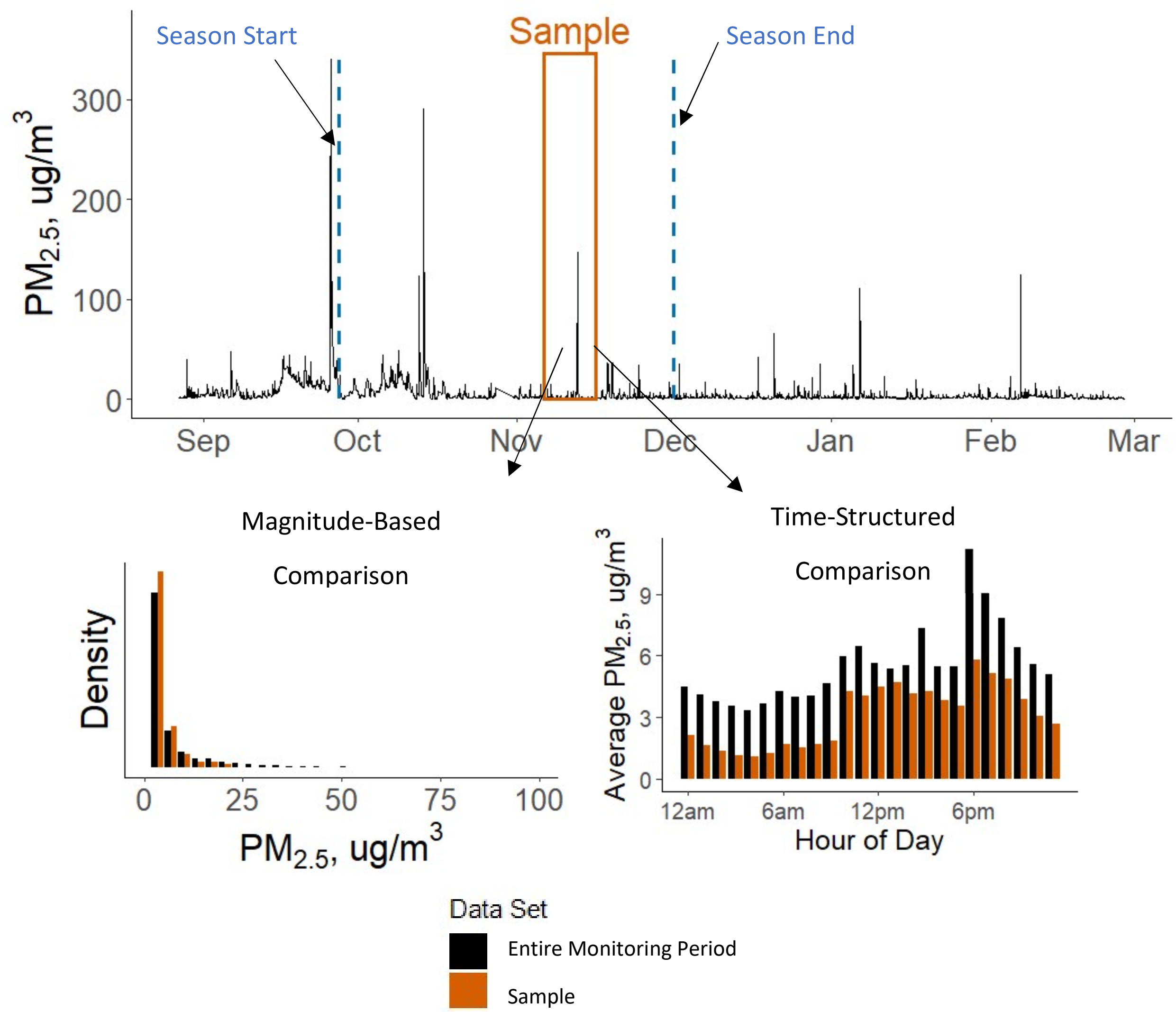


Figure 6: Demonstrations of preparing a $PM_{2.5}$ sample dataset extracted from a season within a long-term monitoring dataset (top) for the calculation of magnitude-based representativeness (bottom left) and time-structured representativeness (bottom right).

When evaluating temporal representativeness of samples in this paper, only time-structured representativeness was analyzed. The hour-of-day average values for a given IEQ metric over a sampling period were used as the sample distribution. The hour-of-day average values for the same IEQ metric over the long-term monitoring period from which the sample was extracted were used as the long-term distribution. Essentially, the discrete method of relative entropy calculation (Equation 4) was used across the 24, one-hour bins in each distribution. This method of time-structured representativeness measures how well the diel structure of a sample characterizes that of the monitoring period from which the sample was extracted.

Magnitude-based representativeness was explored but not included in this study. After difficulties arose with empty distribution bins when using the discrete form of the relative entropy calculation (Equation 4), the discrete method was discarded as an option for magnitude-based calculations. Relative entropy cannot be calculated if the population distribution (the distribution of the entire monitoring period's dataset, in this case) has a value of zero for any of its bins. Zero values were common in distributions for datasets that had wide ranges of values (e.g., from 0 $\mu g/m^3$ to 300 $\mu g/m^3$ for $PM_{2.5}$ datasets). There were also concerns of the effect that arbitrarily chosen bin size can have on the resulting relative entropy values. We chose a bin width of one hour for the distributions used to calculate time-structured representativeness, as hourly intervals are an intuitive averaging period for diel analysis. There was less of a basis for choosing bin width for magnitude-based representativeness. For instance, a $PM_{2.5}$ dataset that contains values between 0 $\mu g/m^3$ and 400 can be broken into 400 bins that are each 1 $\mu g/m^3$ in width. The same sample can be broken into 4000 bins that are each 0.1 $\mu g/m^3$ in width. Bin width was observed to impact the relative entropy values calculated for samples during analysis. Magnitude-based representativeness could have been calculated using the continuous form of relative entropy (Equation 3) to overcome difficulties associated with aggregating data into bins. However, evaluating the continuous form of relative entropy to calculate the representativeness of a sample requires assuming the sample and its respective long-term monitoring period each fit an integrable distribution. A log-normal distribution is an integrable distribution to which IEQ data can most often be fitted; however, this assumption of log-normality would be difficult to test for all samples of lengths ranging from one to 28 days. This assumption also would likely not hold true for many short (e.g., one-day) samples.

IEQ samples that are representative of their respective long-term monitoring period (with respect to diel structure) could be helpful in identifying potential sources of poor IEQ. This is especially true if representative IEQ samples are paired with diel energy data or prior knowledge of residential energy use habits and behaviors. In turn, insight about IEQ and possible sources of poor IEQ could inform what recommendations are made related to improving home energy efficiency. Conversely, if IEQ samples are not representative of long-term conditions, inaccurate conclusions (if any) may be reached, and the ability of energy efficiency programs to leverage home performance upgrades for residential health co-benefits may suffer as a result.

4.3.8.4. Spatial Representativeness and Specificity

Spatial representativeness was also evaluated for the metrics gathered in each room within each home. We used the method similar to the that which was introduced by Osses et al. (2013). They looked at the case of relative entropy between two normally distributed distributions, $a \sim N(\mu_a, \Sigma_a)$ and $b \sim N(\mu_b, \Sigma_b)$, with means μ_a and μ_b and invertible covariance matrices Σ_a and Σ_b. In this normal case, the following relationship holds true:

Equation 6

$$\mathrm{D(a||b)} = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma_b^{-1}\Sigma_a\right) - \mathrm{nm} - \ln\frac{|\Sigma_a|}{|\Sigma_b|} + \Sigma^{-1}(\mu_a - \mu_b)^2\right.$$

Where *tr*(*matrix*), | *matrix* |, and (*matrix*)$^{-1}$ denote the trace, determinant, and inverse of a matrix, respectively; *n* is the number of stations (in our case, three rooms); *m* is the number of species (in our case, one species, as we explore the univariate case). Spatial representativeness was then explored by calculating the "information gain" for each room. Information gain is a measure of decrease in entropy. When used by Osses et al. (2013) and the current study, information gain measures the decrease in entropy (essentially the decrease in uncertainty) that occurs when a new dataset is added to an existing dataset. Assuming the normal case described in Equation 6, information gain can be calculated with the following equation:

Equation 7

$$\text{Information Gain}, \mathrm{I_G^i} = \mathrm{D}(\mathrm{p_{all}}||\mathrm{q_{i\_missing}}) \approx -\frac{1}{2}\ln\frac{|\Sigma_{\mathrm{all}}|}{|\Sigma_{\mathrm{i\_missing}}| * \sigma^2_{\mathrm{i\_missing}}}$$

Where, within the considered home and season, I^i_G is the information gain for room I; p_{all} is the joint probability distribution of values created with data from all three rooms (bedroom, living room, and kitchen); $q_{i\_missing}$ is the probability distribution created via model if measurements from room i are not included. The dataset that represented p_{all}, and which was used to calculate Σ_{all}, was a t x 3 matrix. The three columns in p_{all} were the time-ordered data from each respective room, with t hours of data in each column. In our case, the modeled values of the probability distribution, $q_{i\_missing}$, would be directly dependent on the values of the two "other" (non-i) rooms in the considered home, as no other variables were included in these datasets. Therefore, the dataset that represented $q_{i\_missing}$, and which was used to calculate $\Sigma_{i\_missing}$, was a time-ordered t x 2 matrix; the two columns were the data from the non-i rooms, with t hours of data in each column. Hours when one of the rooms was missing a reading were not considered in the correlation calculation. $\sigma^2_{i\_missing}$ is the summed variance of the datasets from the two other rooms. As the two non-i datasets were likely dependent, the covariance of the two datasets was included when calculating this value.

For this method of measuring information gain, the relative entropy of these two distributions was represented by the third term in Equation 6; this third term is the classical definition of "entropy decrease," (Osses et al., 2013). In this sense, I^i_G essentially measures the reduction in uncertainty that occurs when data from room i is included in the overall dataset. The below equation was then used to calculate a spatial representativeness index value for room i within a given home:

Equation 8

$$\text{Spatial Representativeness}, \mathrm{R_S^i} = \frac{\mathrm{I_G^i} - \mathrm{I_G^{min}}}{\mathrm{I_G^{max}} - \mathrm{I_G^{min}}}$$

Where I_G^{min} and I_G^{max} are the minimum and maximum values of information gain across all three rooms and all seasons within a given home. This method of scaling (subtracting the minimum from both the numerator and denominator) allowed spatial representativeness to be evaluated on a scale of 0 to 1, with larger values implying rooms that are more representative of their respective home. Other scaling methods are possible, but we were more concerned with the resulting order of R^i_S values than their absolute magnitude (i.e., how rooms compared to one another with respect to their representativeness in the home), as there is not an established threshold value for a significant difference in representativeness.

A related value, which we define as "spatial specificity," is the extent to which data gathered from a single sampling location captures trends and measurement magnitudes that are not captured by samples gathered from other sampling locations within the larger location of interest. We evaluated spatial specificity via the entropy-related measure of mutual information (Equation 9):

Equation 9

$$\text{Mutual Information}, \mathrm{I_M^i} = \mathrm{D}\left(\mathrm{p_{all}} \middle\| \mathrm{p_i p_{i\_missing}}\right) = -\frac{1}{2}\ln \frac{|\Sigma_{\mathrm{all}}|}{|\Sigma_{\mathrm{i\_missing}}| * \sigma_\mathrm{i}^2}$$

Where p_i is the marginal probability distribution created from the dataset of room *I*; $p_{i\_missing}$ is the marginal probability distribution created from the dataset of the two non-*i* rooms; σ_i^2 is the variance of the room *i* dataset. Osses et. al (2013) defined this as specificity, while others (Perez-Abreu & Rodriguez, 1996; Silva & Quiroz, 2003) defined this as "effectiveness." Regardless, mutual information, when used in this sense, is essentially a measure of how difficult it is to produce data observed within a considered room if data from that room is not available. Equation 10 was then used to calculate a spatial specificity index value for room *i* within a given home:

Equation 10

$$\text{Spatial Specificity}, \mathrm{S_S^i} = 1 - \frac{\mathrm{I_M^i}}{\mathrm{I_M^{max}}}$$

Where I_M^{max} is the maximum value of mutual information across all rooms and seasons within a given home. This scaling method allows spatial specificity to be evaluated on a scale of 0 to 1, with larger values implying rooms that are more specific. Similar to representativeness, a different scaling method could have been chosen, but we are more interested in the order of the resulting specificity values between rooms within a home than we are of the magnitude of these values.

These spatial measures of representativeness and specificity essentially consider the entire time-structure *and* magnitudes of time series data values. This is because both measures consider the variance of each dataset and the correlation between time-ordered values within datasets in each home and. This correlation is possible with spatial representativeness and specificity calculations because the datasets of rooms within the same home can be paired. The sample and long-term datasets used for temporal representativeness calculations could not be paired due to their different sizes/time periods. The measures of spatial representativeness and spatial specificity were each calculated separately for each season within each home to test whether these measures varied between seasons. Only homes with data from all three rooms within a given season were used in spatial analysis for the given season. The assumption of normality was required for the use of Equation 7 and Equation 9. The datasets of each IEQ indicator collected from each home-room-season condition therefore had to be checked for normality before spatial specificity and spatial representativeness of the datasets were calculated. If a dataset did not to satisfy the assumption of normality, the dataset had to be transformed to fit a normal distribution prior to representativeness or specificity calculation.

Table 5: Methods of evaluating representativeness and specificity of metrics.

Quality Criterion		Definition	Method of Evaluation
Temporal Representativeness	Time-Structured	Extent to which sampling period captures diel trends that characterize average conditions in long-term period	Relative Entropy $D_s = \sum_{x \in A} p(x) \ln \frac{p(x)}{q(x)}$
	Magnitude-Based	Extent to which sampling period captures magnitudes that characterize average conditions in long-term period	Representativeness $R_T^s = 1 - \frac{D_s}{D_{max}}$
Spatial Specificity		Extent to which data recorded in a sampling location captures trends and magnitudes that are not captured by samples recorded in other sampling locations within the larger location of interest	Mutual Information $I_M^i = -\frac{1}{2} \ln \frac{\lvert\Sigma_{all}\rvert}{\lvert\Sigma_{i\_missing}\rvert * \sigma_i^2}$ Specificity $S_S^i = 1 - \frac{I_M^i}{I_M^{max}}$
Spatial Representativeness		Extent to which data recorded from a sampling location captures trends and magnitudes that characterize average conditions within the larger location of interest	Information Gain $I_G^i \approx -\frac{1}{2} \ln \frac{\lvert\Sigma_{all}\rvert}{\lvert\Sigma_{i\_missing}\rvert * \sigma_{i\_missing}^2}$ Representativeness $R_S^i = \frac{I_G^i - I_G^{min}}{I_G^{max} - I_G^{min}}$

5. Results

5.1. Study Population

Sixteen homes were enrolled into the study. Home 5 was omitted from all analysis, as less than two weeks of data were collected from this home. This resulted in an analyzed study population of fifteen homes. Table 6 displays a summary of characteristics of homes included in analysis. Two thirds of the recruited homes were owner-occupied, and the rest were renter-occupied. All owner-occupied homes were detached single-family homes (i.e., they did not share a common wall with any other dwellings). Four of the renter-occupied homes were duplex units. Each duplex unit was a single-family attached home (i.e., shared a common wall with another duplex unit). All duplex units were considered individual homes for the purposes of this study. The remaining renter-occupied unit was a detached single-family home. No socio-demographic or income data were available for individual homes, as these data were not gathered in the energy assessment or participant surveys. Most of the recruited homes were in census tracts for which the median household income was between $50,000 and $89,999. Fort Collins is situated within Larimer County, Colorado. The median household income for Larimer County in 2019 was estimated at $75,186 (United States Census Bureau, 2019). Broad ranges of conditioned area (total amount of floor space in which thermal conditions are controlled), number of occupants, and age were observed in the recruited homes. Most of the recruited homes contained two stories and had attached garages.

Six to ten months of five-minute resolution IEQ data were collected in each enrolled home between July 20, 2020 and May 31, 2021. Start dates were rolling, and ranged from July 20, 2020 to November 23, 2020. The number of hours for which valid IEQ sensor data were recorded indoors ranged from 3,248 hours (136 days) to 7,507 hours (315 days), with a median of 6,589 hours (275 days) across all homes. A room-hour is defined as one hour of data in one room. Up to three IEQ sensors were placed indoors in each home, excluding outdoor and garage sensors. Between 9,744 and 21,690 valid room-hours, with a median of 18,926 room-hours were recorded by indoor sensors across all homes. IEQ sensors occasionally malfunctioned due to battery loss or Wi-Fi disconnection. All fifteen homes recorded valid IEQ data over at least 91% of their corresponding enrollment period; for most homes, this percentage was above 98%. Two rounds of filter-based $PM_{2.5}$ calibration visits were conducted for most homes within the study period. Malfunctioning of filter-based $PM_{2.5}$ sampling devices resulted in loss of filter data (28% of filter samples). Due to home energy monitor data import issues, two homes completely lacked energy use data and four homes lacked data after mid-May of 2021. One home lacked data after late March 2021 due to Wi-Fi connection issues.

*Table 6: Summary of homes enrolled in study and included in analysis. Information gathered from questionnaire and energy assessment data. *Data not available from survey or assessment. **Includes two sets of duplex units and one single-family detached home. ***Median annual household income for census tract in which home was located. ****Air changes per hour for home pressurized at 50 psi, reported in home energy assessment.*

Variable	Number of Homes	Percent of Homes
Year Built		
1900-1920	1	7%
1960-1979	8	53%
1980-1999	3	20%
2000-2020	3	20%
Conditioned Area, sq. ft.		
<1499	5	33%
1500-2499	3	20%
2500-3499	6	40%
>3500	1	7%
Garage Type		
Attached	12	80%
Detached	3	20%
Floors		
1	2	13%
2	12	80%
3	1	7%

*Table 6 cont.: Summary of homes enrolled in study and included in analysis. Information gathered from questionnaire and energy assessment data. *Data not available from survey or assessment. **Includes two sets of duplex units and one single-family detached home. ***Median annual household income for census tract in which home was located. ****Air changes per hour for home pressurized at 50 psi, reported in home energy assessment.*

Variable	Number of Homes	Percent of Homes
Stove Type		
Electric	8	53%
Natural Gas	4	27%
Unknown*	3	20%
Property Type		
Owner-Occupied	10	67%
Renter-Occupied	5**	33%
Occupants		
1	1	7%
2	7	47%
3	1	7%
4	3	20%
5	2	13%
Unknown*	1	7%
Census Tract Median Annual Household Income* **		
\$20,000 - \$49,999	1	7%
\$50,000 - \$89,999	9	60%
\$90,000 - \$119,999	4	27%
\$120,000 - \$149,999	1	7%
ACH_{50}****		
≤ 3.0	1	7%
3.1 to 6.0	4	27%
6.1 to 9.0	5	33%
> 9.0	2	13%
Unknown*	3	20%

5.2. Calibration of Real-Time $PM_{2.5}$ Measurements

Ninety-six (96) pairs of filter-based and sensor-based $PM_{2.5}$ measurements were collected over the study period. Three of these pairs were omitted from analysis because the filter-based measurement resulted in a negative average concentration. Negative average concentrations could have occurred due to the presence of residue on the filter during pre-deployment weighing, and the loss of this residue during deployment. If less mass was collected on the filter over the sample collection period than the mass of the initial residue, negative concentrations could have been calculated post-deployment. We evaluated the relationship between the filter- and sensor-based measures of $PM_{2.5}$ in two ways (Figure 7). Pairwise ratios between the gravimetric $PM_{2.5}$ mass concentration and the time-weighted $PM_{2.5}$ concentration averaged over the same sampling period as the filter deployment ranged from 0.09 to 5.21, with a median of 0.88. Alternatively, a simple linear regression of sensor-based, time-weighted average $PM_{2.5}$ concentrations versus gravimetric $PM_{2.5}$ mass concentrations yielded a slope of 0.29, when no data were excluded (Table 7). Excluding one pairwise sample (1.1% of the total) that was not within 2.5 standard deviations from the mean, the regression coefficient was 0.28. For the purpose of this study, we used the pairwise correction factor – i.e., the ratio of the gravimetric $PM_{2.5}$ mass concentration and the time-weighted $PM_{2.5}$ concentration – to adjust the time-weighted $PM_{2.5}$ concentrations for the corresponding sensor.

Table 7: Results of simple model (TWA = EstimateC_{filter} + 0) with no data omitted. Excluding one pairwise sample (1.1% of the total) that was not within 2.5 standard deviations from the mean, the regression estimate = 0.28*

Term	Estimate	p-value	95% Confidence Interval
C_{filter}	0.29	0.00	0.20 - 0.37

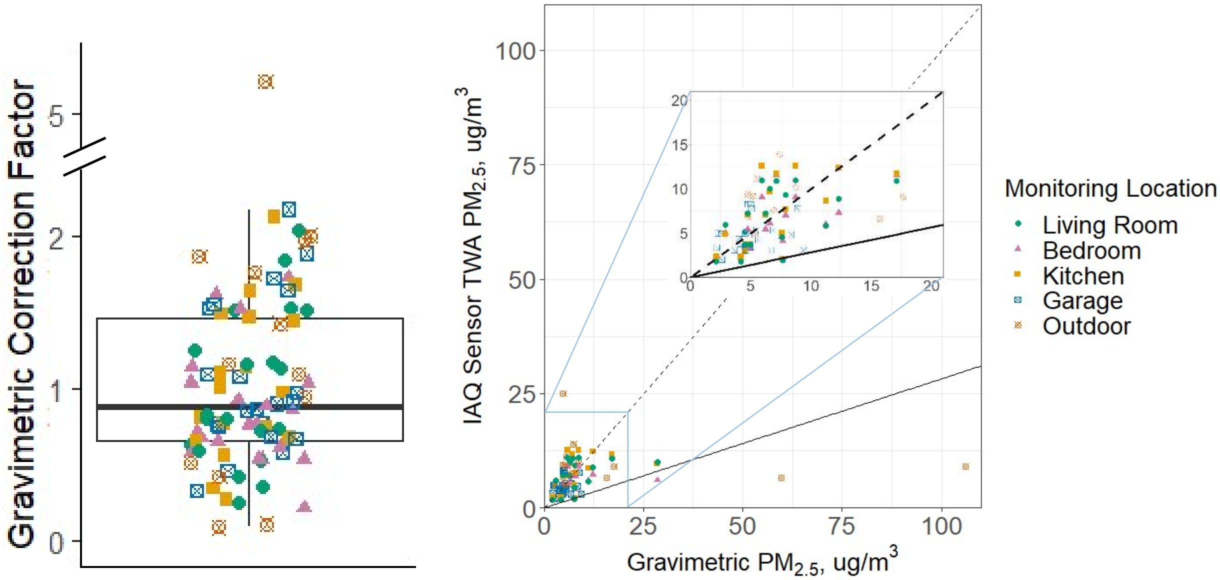


Figure 7: Left: Distribution of gravimetric correction factors calculated from paired IAQ sensor $PM_{2.5}$ TWA values and filter-based average (Gravimetric) $PM_{2.5}$ values. Center: IAQ sensor TWA $PM_{2.5}$ values vs filter-based average (Gravimetric) $PM_{2.5}$ values (C_{filter}). Dotted line shows theoretical 1 to 1 line; solid line shows simple model results plotted. Inset plot included for view of smaller values.

5.3. Comparison of Home Energy Monitor Measurements with Natural Gas Heating Data

Natural gas usage data were collected in a subset of homes and compared with our measures of heating device usage derived from the home energy monitor usage (methodology described in Section 4.3.6). The amount of natural gas consumed by each home was available by monthly billing period. Monthly natural gas consumption for each home was normalized to the month with the highest gas consumption. The amount of time on for the identified primary heating device was calculated for the same monthly billing periods within each home. These monthly time on values were also normalized and paired by billing month to the normalized gas usage data (Figure 8). No detailed analysis between energy monitor data and natural gas consumption data was performed; however, the general trend of the monthly natural gas data agreed with the general trend of the monthly primary heating device usage reported by the home energy monitor. Generally, as monthly natural gas usage increased, so did the amount of time on for the primary heating device. Certain homes displayed uncertainty in the method used to identify a primary heating device with the home energy monitor. For example, some billing months (October and November) in Homes 9, 10, and 11 displayed little usage for the primary heating device, while natural gas usage was substantial in these months (Figure 8). The home energy monitor may not have identified the primary heating device during these months even though the heating device was running. This discrepancy between gas consumption and heating device usage suggests the primary heating device usage methodology was not an effective proxy for home heating intensity for all homes. However, the generally acceptable agreement observed between the heating device and natural gas usage for most of the homes in this subset suggests the use of the primary heating device proxy is justified for defining seasons.

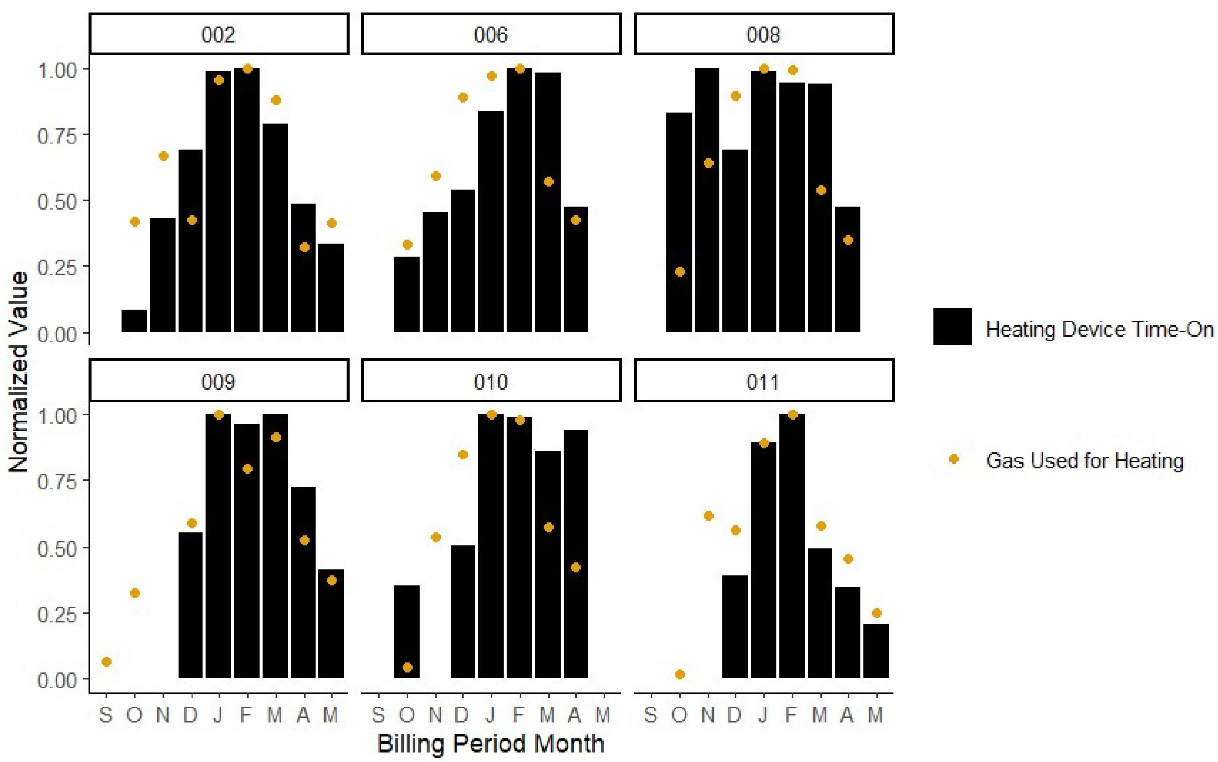


Figure 8: Amount of time-on of heating device, as measured by energy monitor device, and amount of gas used for heating, as measured by home natural gas supplier, during billing month (noted on x-axis) for select homes (home number above each respective plot). Billing months span from September 2020 (S) to May 2021 (M). Both time-on and gas usage numbers were normalized so that the month with greatest magnitude within each home was assigned a value of 1.

5.4. Autocorrelation of IEQ Metrics

Autocorrelation analysis was performed using the methodology described previously (Section 4.3.3). The number of days required before autocorrelation became insignificant (when data was no longer considered autocorrelated) was determined for $PM_{2.5}$, TVOC, CO_2, and temperature datasets from each indoor sensor in each season (Figure 9). The median of, and variation in, the length of time (i.e., number of days) until autocorrelation became statistically insignificant was comparable between indoor samples of TVOC, CO_2, and temperature. Autocorrelation was generally present in indoor samples of these three IEQ indicators longer than autocorrelation was present in indoor samples of $PM_{2.5}$.

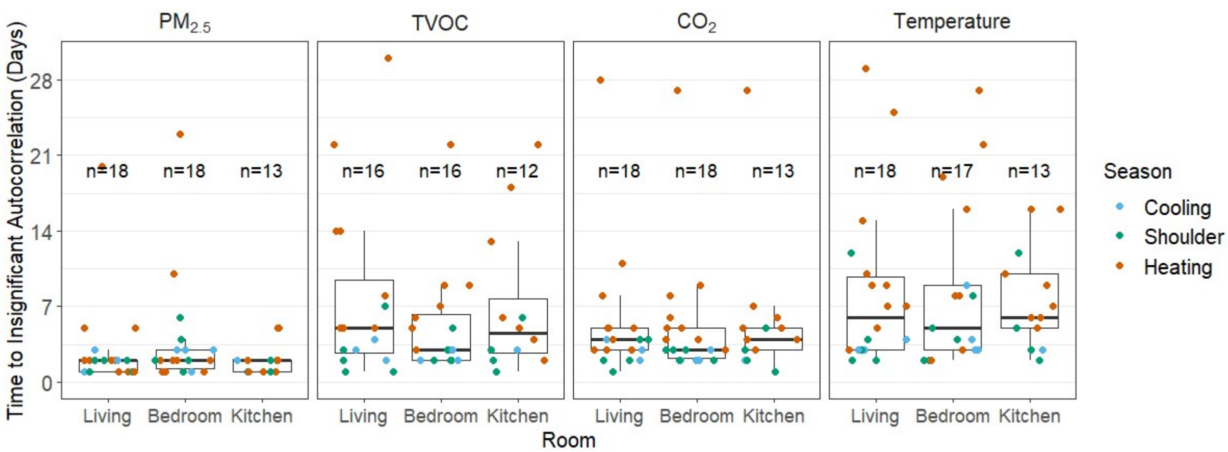


Figure 9: Number of days (in lagged correlations) before insignificant autocorrelation is reached for each IEQ indicator, homes pooled, stratified by room and colored by season. Room-season samples consisting of less than 25 days of data, missing more than 11.1% (1/9) of monitored days, or that did not reach insignificance prior to 30 days were omitted from analysis.

Autocorrelation results determined the averaging period required to achieve independent samples for each IEQ indicator. Data recorded during the heating season appeared to remain autocorrelated longer, on average, than the other two seasons (Figure 9), although no analysis was performed to test for a significant difference in autocorrelation between the seasons. We assumed averaging period length for all seasons for each IEQ indicator. For each IEQ indicator, the median number of days before data was no longer autocorrelated (pooling datasets from all seasons) was used as the averaging period required for an independent sample (Table 8). The resulting averaging period for indoor $PM_{2.5}$ samples (two days) was the shortest of the four IEQ indicators. Averaging periods of indoor CO_2, TVOC, and temperature samples were longer, at four, five and six days respectively.

*Table 8: Averaging period for independent samples for each IEQ metric, determined using the median time (in days) at which the autocorrelation factor in autocorrelation analysis (pooled homes, seasons, and indoor rooms) reached insignificance. *Rounded up to nearest integer day value.*

IEQ Indicator	Averaging Period (days)
$PM_{2.5}$	2
TVOC	5*
CO_2	4
Temperature	6

5.5. Summary Statistics of IEQ Indicators

5.5.1. Overall and Seasonal Averages

Time-weighted hourly average values were calculated from the five-minute measurements recorded from each indoor sensor. These values were then pooled across all rooms (kitchen, living room, and bedroom) within each home to summarize the data collected over the entire study period (Table 9

). The US Environmental Protection (EPA) does not provide guidelines for levels of air quality pollutants indoors, but it has set standards for criteria pollutants in ambient (outdoor) air for both short-term and long-term exposure (US Environmental Protection Agency, 2021). The long-term EPA primary standard ("long-term" is defined as the annual average over three years; "primary" pertains to the protection of sensitive populations) is 12 μg/m^3. The World Health Organization (WHO) has set air quality guideline (AQG) levels for short-term and long-term exposure to $PM_{2.5}$, which the organization stated were relevant to both indoor and outdoor exposures (2021). The WHO recently updated their guidance for annual $PM_{2.5}$ levels, such that the air quality guidance (AQG) level is now 5 μg/m^3. The time-weighted average across all 15 homes in the study (8.2 μg/m^3) was below the EPA standard but exceeded the WHO ACG. Some homes also exceeded the EPA standard (Table 9).

The EPA has not established guidelines or standards for airborne volatile organic compounds (US Environmental Protection Agency, n.d.), and the WHO has only suggested guidelines for specific organic compounds (e.g., benzene, naphthalene, and formaldehyde), not aggregate compound levels which are measured with the TVOC metric (World Health Organization, 2010). LEED, a green building certification program that is prevalent worldwide, has set a maximum TVOC concentration guideline of 500 μg/m^3 as a threshold for a building to achieve a credit among enhanced air quality and enhanced indoor environmental quality criteria (U.S. Green Building Council, 2018). The time-averaging period for this guideline is unspecified, but LEED guides refer to samples taking place in a single day during building occupied hours. As such, 500 μg/m^3 TVOC concentrations have been used as a level for comparison in scientific IEQ literature (Jia et al., 2019; U.S. Green Building Council, 2018). The response of a TVOC sensor does not provide any information on the mixture of VOCs to which the sensor may be responding. Therefore, converting a mass concentration to a part per billion by volume concentration (i.e., ppb) requires making an assumption about the molar mass of a hypothetical mixture of VOCs. We assumed the TVOC sensor in our study was responding to a mixture of 22 VOCs with ratios defined by Sensirion (2019) and Mølhave et al. (1997). This assumption implies a concentration of 1 ppb corresponds to 4.5 μg/m^3, meaning the LEED TVOC guideline converts to approximately 110 ppb. The average indoor TVOC concentration in each home over the study period exceeded this 110 ppb guideline, and the pooled average of all homes was more than triple this guideline.

Table 9: Time-weighted hourly average values calculated from the five-minute measurements recorded from all indoor sensors (kitchen, living room, and bedroom), all rooms pooled within each home. The time-weighted hourly average values for all homes at the bottom of the table were calculated pooling data from all indoor rooms from all homes. Data is from entire study period for each home.

Home	N (room-hours)	$PM_{2.5}$ (ug/m^3)			TVOC (ppb)			CO_2 (ppm)		
		Range	Mean ± SD	Median	Range	Mean ± SD	Median	Range	Mean ± SD	Median
1	11,544	0.0 - 183.4	3.6 ± 6.9	1.7	21 - 3,896	255 ± 175	234	400 - 2,973	763 ± 340	682
2	15,005	0.0 - 443.0	5.1 ± 11.7	2.5	23 - 5,943	335 ± 334	251	400 - 4,014	736 ± 449	587
3	21,690	0.0 - 806.5	13.8 ± 30.9	4.9	20 - 10,018	509 ± 507	366	400 - 5,000	1,004 ± 656	804
4	19,525	0.0 - 865.8	10.5 ± 26.5	2.1	20 - 6,991	217 ± 247	165	400 - 1,997	639 ± 183	605
6	20,213	0.0 - 405.0	4.2 ± 9.7	1.9	20 - 9,690	300 ± 239	261	400 - 1,802	620 ± 167	581
7	19,191	0.0 - 421.6	6.5 ± 14.3	2.3	20 - 6,542	323 ± 300	255	400 - 1,581	637 ± 166	601
8	20,086	0.0 - 451.7	9.7 ± 18.0	4.8	20 - 14,068	488 ± 512	338	400 - 2,034	666 ± 219	613
9	19,396	0.0 - 771.2	7.9 ± 17.3	3.8	9 - 9,414	565 ± 525	382	400 - 2,883	952 ± 377	898
10	19,120	0.0 - 515.4	7.7 ± 18.6	3.4	20 - 14,337	176 ± 262	107	399 - 2,449	753 ± 309	674
11	14,967	0.0 - 1,131.6	14.0 ± 36.2	5.6	20 - 6,746	412 ± 425	302	400 - 2,206	854 ± 282	812
12	18,926	0.0 - 1,066.7	4.0 ± 12.6	1.5	21 - 3,887	218 ± 127	200	400 - 1,929	622 ± 218	558
13	16,347	0.0 - 1,121.6	11.7 ± 41.1	3.8	20 - 7,311	318 ± 282	261	400 - 3,275	715 ± 342	623
14	14,912	0.0 - 90.8	2.4 ± 3.0	1.6	20 - 3,535	255 ± 170	229	400 - 3,641	704 ± 294	637
15	11,445	0.0 - 1,131.6	16.3 ± 74.4	2.9	0 - 27,984	406 ± 563	305	333 - 3,586	823 ± 476	691
16	9,744	0.0 - 756.6	5.3 ± 20.9	2.5	20 - 2,961	216 ± 158	197	400 - 2,022	770 ± 277	721
All homes	252,111	0.0 - 1,131.6	8.2 ± 27.0	2.8	0.0 - 27,984	340 ± 377	252	333 - 5,000	749 ± 364	645

Table 9 cont.: Time-weighted hourly average values calculated from the five-minute measurements recorded from all indoor sensors (kitchen, living room, and bedroom), all rooms pooled within each home. The time-weighted hourly average values for all homes at the bottom of the table were calculated pooling data from all indoor rooms from all homes. Data is from entire study period for each home.

Home	N (room-hours)	Temperature (^oC)			Relative Humidity (%)		
		Range	Mean ± SD	Median	Range	Mean ± SD	Median
1	11,544	12.5 - 28.1	21.0 ± 2.3	21.0	12.9 - 55.3	34.7 ± 5.7	34.7
2	15,005	13.5 - 26.3	20.9 ± 1.7	20.9	8.0 - 64.1	31.1 ± 11.4	28.4
3	21,690	13.3 - 29.2	20.5 ± 2.1	20.2	11.7 - 87.0	43.5 ± 7.8	42.8
4	19,525	14.5 - 29.5	18.8 ± 2.5	18.3	11.0 - 69.7	37.1 ± 6.9	36.6
6	20,213	18.3 - 29.9	20.9 ± 1.3	20.5	16.1 - 62.4	43.9 ± 5.7	44.9
7	19,191	13.3 - 28.6	20.7 ± 2.0	20.7	9.7 - 68.6	35.8 ± 6.9	35.7
8	20,086	-3.2 - 36.9	19.4 ± 4.2	20.8	9.5 - 67.1	35.2 ± 7.4	33.7
9	19,396	11.4 - 31.3	19.5 ± 2.5	19.4	10.4 - 70.1	42.1 ± 6.4	42.0
10	19,120	9.9 - 28.7	19.9 ± 2.2	19.8	7.9 - 67.9	33.0 ± 9.1	31.9
11	14,967	15.1 - 34.3	20.2 ± 2.1	19.8	12.1 - 72.2	42.7 ± 4.9	42.8
12	18,926	10.7 - 25.8	20.2 ± 2.0	20.6	14.0 - 57.2	34.9 ± 5.6	34.0
13	16,347	10.2 - 26.9	19.5 ± 2.4	19.2	9.1 - 62.5	34.0 ± 7.8	33.3
14	14,912	15.8 - 26.7	21.0 ± 1.3	21.0	11.1 - 55.6	26.5 ± 6.9	25.0
15	11,445	8.0 - 27.3	18.5 ± 2.6	18.6	15.3 - 64.4	38.0 ± 6.5	37.4
16	9,744	9.3 - 30.3	18.0 ± 2.7	17.8	11.2 - 56.2	31.4 ± 6.3	30.8
All homes	252,111	-3.2 - 36.9	20.0 ± 2.5	20.2	7.9 - 87.0	36.7 ± 8.7	37.0

To evaluate median indoor values for concentrations of $PM_{2.5}$, TVOC, and CO_2 in this study, hourly concentrations were averaged to the sample length corresponding to the minimum length determined to support independence of samples from the autocorrelation analysis. These averages were calculated for each indoor room in all homes, disaggregated by season, then pooled within each season for analysis (Figure 10 and Figure 11). Overall, median $PM_{2.5}$, TVOC, and CO_2 concentrations were lower in the heating season than in the cooling and shoulder seasons. Ranges of indoor $PM_{2.5}$, TVOC, and CO_2 concentrations were comparable between seasons. Median indoor temperature was 1 to 2 degrees (Celcius) lower in shoulder season days compared to those of cooling season days, and the same was true for heating season days with respect to shoulder season days. The same analysis was performed disaggragated by room type, with less noticable differences displayed between rooms (Figure 30 and Figure 31 in Appendix).

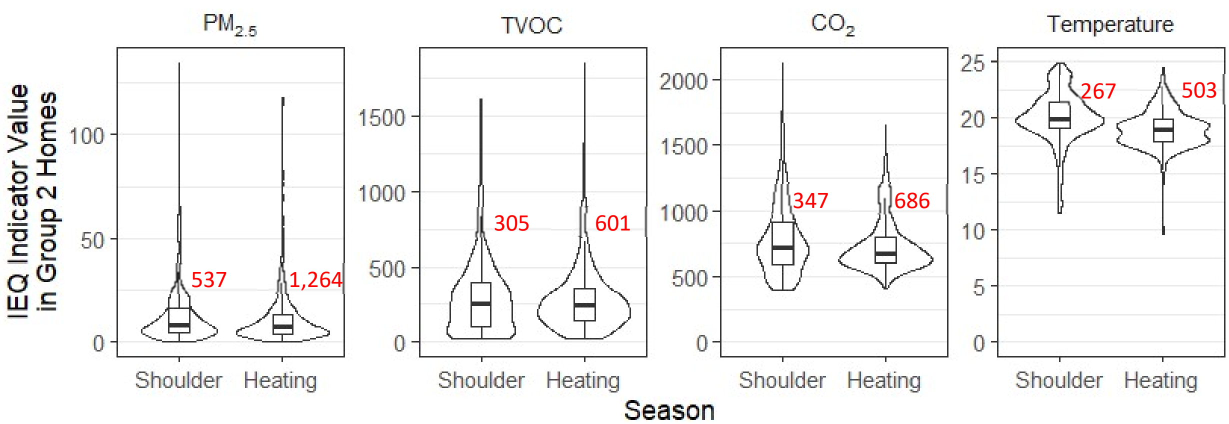


Figure 10: Distributions of time-averaged hourly IEQ parameter values from indoor rooms in Group 2 Homes (homes with no definable cooling period), shown separately by season. Values from all rooms in all Group 2 homes pooled. Units: $PM_{2.5}$ (µg/m³), TVOC (ppb), CO_2 (ppm), Temperature (ºC). Red value shows number of independent samples included in each distribution

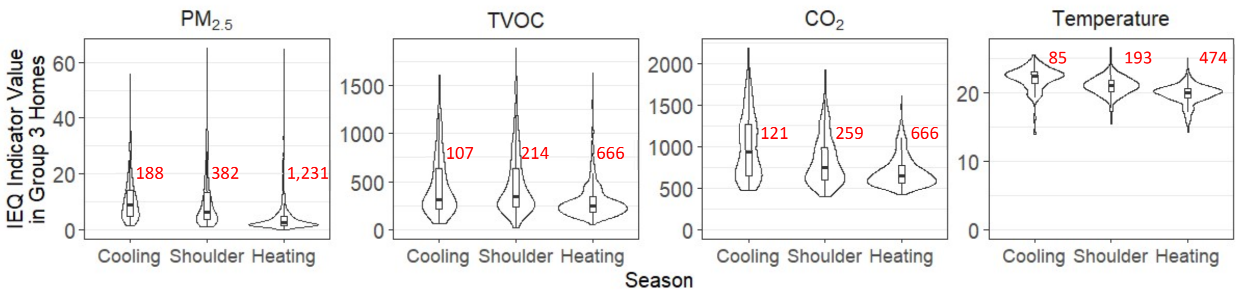


Figure 11: Distributions of time-averaged hourly IEQ parameter values from indoor rooms in Group 3 Homes (homes with definable cooling periods), shown separately by season. Values from all rooms in all Group 3 homes pooled. Units: $PM_{2.5}$ (μg/m³), TVOC (ppb), CO_2 (ppm), Temperature (°C). Red value shows number of independent samples included in each distribution

5.5.2. Coefficient of Variation

Coefficients of variation were calculated for each IEQ indicator over the entire study period for each home (Table 10). Time-weighted hourly average values were first calculated from the five-minute measurements recorded from each indoor sensor. Coefficients of variation were calculated from these values, pooling across all rooms (kitchen, living room, and bedroom) within each home. $PM_{2.5}$ varied most with respect to its average within homes, followed by TVOC and CO_2. Coefficients of variation were plotted and stratified by indoor room and season separately for Group 2 homes (Figure 12) and Group 3 homes (Figure 13). No clear difference between coefficient of variation was observed between rooms or between seasons, although the median value of coefficient of variation for indoor rooms in the heating season was larger than those for the cooling and shoulder seasons in Group 3 homes.

Table 10: Coefficient of variation calculated for each metric over the entire study period for each home. Time-weighted hourly average values were first calculated from the five-minute measurements recorded from each indoor sensor (kitchen, living room, and bedroom). Coefficients of variation were calculated from these values, pooling across all rooms within each home. Values for all homes at the bottom of the table were calculated pooling data from all indoor rooms from all homes

Home	N (room-hours)	$PM_{2.5}$ (ug/m^3) COV	TVOC (ppb) COV	CO_2 (ppm) COV	Temperature (°C) COV	Relative Humidity (%) COV
1	11,544	1.94	0.69	0.45	0.11	0.16
2	15,005	2.30	1.00	0.61	0.08	0.37
3	21,690	2.24	1.00	0.65	0.10	0.18
4	19,525	2.53	1.14	0.29	0.13	0.19
6	20,213	2.28	0.80	0.27	0.06	0.13
7	19,191	2.19	0.93	0.26	0.10	0.19
8	20,086	1.86	1.05	0.33	0.22	0.21
9	19,396	2.18	0.93	0.40	0.13	0.15
10	19,120	2.41	1.49	0.41	0.11	0.28
11	14,967	2.58	1.03	0.33	0.10	0.12
12	18,926	3.14	0.58	0.35	0.10	0.16
13	16,347	3.52	0.89	0.48	0.12	0.23
14	14,912	1.26	0.67	0.42	0.06	0.26
15	11,445	4.58	1.38	0.58	0.14	0.17
16	9,744	3.91	0.73	0.36	0.15	0.20
All homes	252,111	3.27	1.11	0.49	0.13	0.24

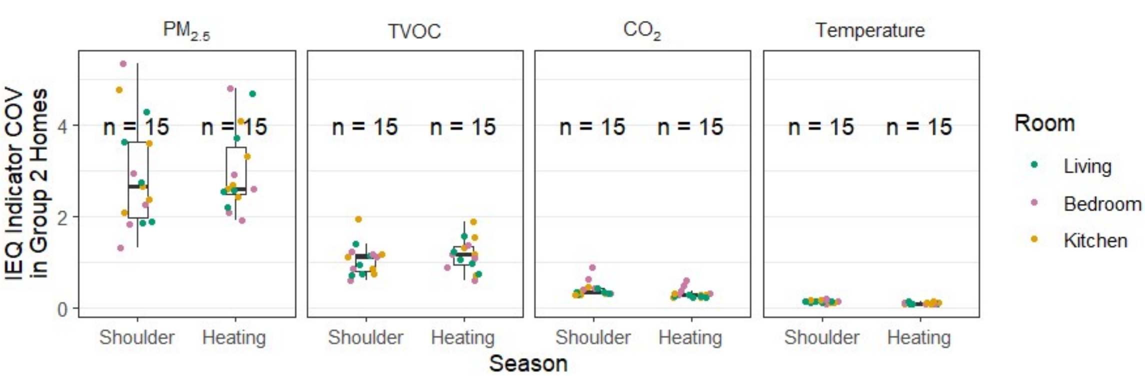


Figure 12: Coefficient of variation for hourly data of IEQ metrics in Group 2 Homes (homes without definable cooling periods), separated by season and colored by sampling room. "n" is the count of home-rooms plotted for each season.

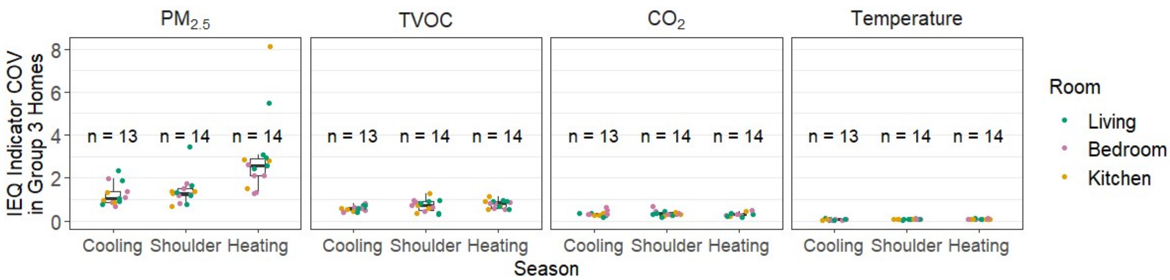


Figure 13: Coefficient of variation for hourly data of IEQ metrics in Group 3 Homes (homes with definable cooling periods), separated by season and colored by sampling room. "n" is the count of home-rooms plotted for each season.

5.6. Diel Trends

5.6.1. Trends by Location

Diel average concentrations of $PM_{2.5}$, TVOC, and CO_2 were calculated with time-weighted hourly average data from all homes pooled together, disaggregated by monitor location (Figure 14). The 24-hour standard for $PM_{2.5}$ (98th percentile of daily concentrations averaged over three years) is 35 µg/m^3. The WHO has set interim targets for air quality guidance (AQG), but the ultimate AQG goal for 24-hour (99th percentile of daily concentrations averaged over one year) $PM_{2.5}$ exposure is 15 µg/m^3. None of the monitored locations reached the EPA's guideline of 35 µg/m^3 for any hour-of-day on average. Levels measured in indoor rooms (kitchen, living room, and bedroom) approached or exceeded the WHO AQG of 15 µg/m^3 during evening hours, but average daily values were lower. Some individual homes exceeded the WHO AQG more consistently throughout the day on average (Figure 32 and Figure 33 in Appendix). In indoor rooms, average $PM_{2.5}$ tended to increase from 6 am until 9 am, plateau until the afternoon, and peak around 6 pm. While outdoor sensors on average did have morning and afternoon peaks for $PM_{2.5}$, the peak concentrations were not as high compared to indoor peaks. Garage $PM_{2.5}$ levels were consistently low, on average, throughout the day.

TVOC concentrations in kitchens and living rooms, on average, showed similar within-day temporal trends as $PM_{2.5}$. TVOC concentrations increased between 6 am and 9 am, plateaued throughout the day, and peaked around 7pm. This evening peak also occurred in garage TVOC concentrations with a consistently higher average magnitude. TVOC concentrations in garages generally decreased through the night until about mid-day. The garage experienced more variability in TVOC concentrations compared to indoor rooms, seen in the wider confidence intervals. This variability appears to arise from between-home differences when the data is disaggregated by home (Figure 32 and Figure 33 in Appendix). As solvents, cleaning products, and fuels are common sources of VOCs, the high between-home variability in TVOC concentration may be attributable to the between-home variability in types and amounts of such products participants store in their garages. The amount of time the garage door is open during a typical day may also have high variability between homes, which is dependent on occupant behavior. Opening and closing of the garage door would likely impact the air exchange rate within the garage, and thus the garage TVOC concentration. Bedrooms, on average, exhibited TVOC concentrations similar in magnitude to kitchens and living rooms, although peaks occurred overnight instead of in the evening. Participant use of personal care products before sleeping could have been the cause for elevated TVOC concentrations near the bedroom. Also, human skin is itself a source of VOCs (Gallagher et al., 2008), which could cause higher TVOC concentrations in the bedroom when participants are preparing to sleep. Again, it should be noted that TVOC is an aggregate measure of volatile organic compounds, and the organic compounds observed may not be the same between locations or between different times of day.

Average diel CO_2 concentrations also had time-of-day peaking trends similar to $PM_{2.5}$ and TVOC in kitchens and living rooms. Peak average bedroom CO_2 concentrations were significantly higher than the other two rooms and occurred overnight. Average outdoor and garage CO_2 concentrations stayed approximately constant at typical background concentrations (400 ppm) throughout the day.

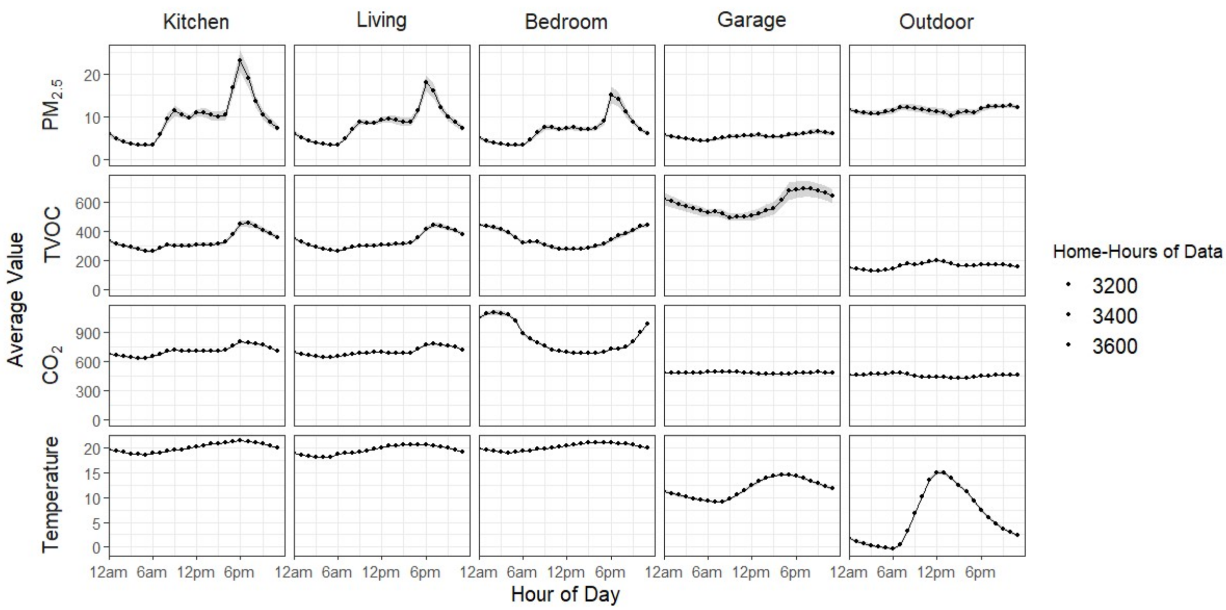


Figure 14: Average hour of day concentrations of $PM_{2.5}$ in µg/m³ (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in ºC (bottom row) calculated from average hourly values over entire monitoring period for all homes pooled together, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean.

5.6.2. Trends by Season

Diel average concentrations of $PM_{2.5}$, TVOC, and CO_2 were disaggregated by season-days and plotted. The classifications of heating, shoulder, and cooling day that had been assigned to each individual day (methodology described in Section 4.3.6) were used to group days together. Data from all living room, bedroom, and kitchen sensors in homes were pooled together for ten out of sixteen homes. Five homes were excluded from this evaluation because data were missing (i.e., homes 3 and 16) or because of ambiguity noted previously (Section 4.3.6) about what the home energy monitor data represented (i.e., homes 6, 8, and 14). Homes that had no cooling days were plotted separately from homes with cooling days.

Average hour-of-day pollutant levels in homes with three seasons were lower for heating days than for the other two seasons. All homes, on average, had more pronounced morning $PM_{2.5}$ peaks during shoulder season days than on heating days. Homes with cooling days exhibited a morning $PM_{2.5}$ peak that, on average on cooling days, was higher than the evening peak. Peak CO_2 levels were also observed to be higher on average in mornings during cooling days compared to the other two seasons.

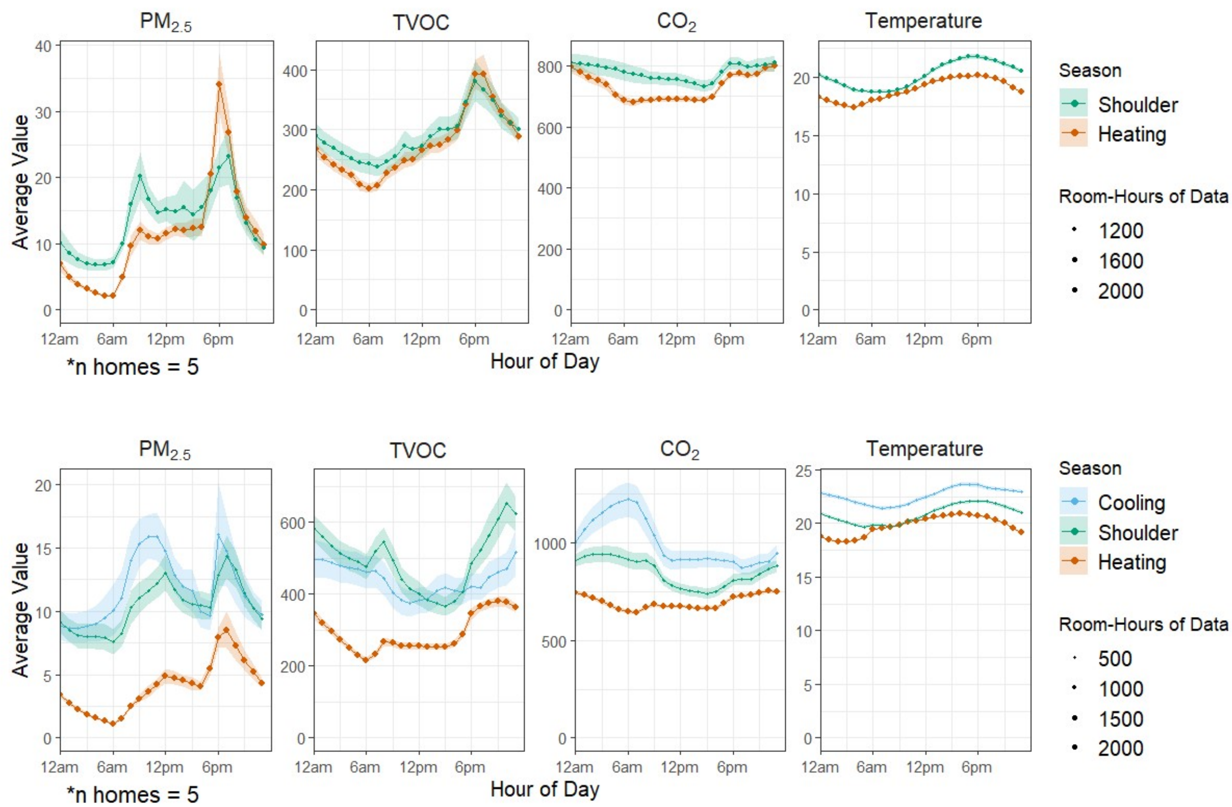


Figure 15: Average hour of day values of $PM_{2.5}$ in µg/m³, TVOC in ppb, CO_2 in ppm, and temperature in ºC calculated from average hourly values for days in each season from kitchen, living room, and bedroom sensors in Group 2 Homes (homes with no definable cooling periods) pooled together (top) and Group 3 Homes (homes with definable cooling periods) pooled together (bottom). Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean. Size of point is proportional to the amount of hourly average values used to calculate an average value for each hour of day.

5.6.3. Correlations Between Rooms

Long-term high-resolution data allowed us to explore if, for a given IEQ metric, the strength of correspondence between hourly data in two locations within a home was affected by the hour of the day. Pearson correlations were calculated by hour-of-day for $PM_{2.5}$ concentrations over the entire monitoring period, separately for each home (Figure 16).

Living room $PM_{2.5}$ concentrations (which may be the room, or one of the rooms, where families spent the highest proportion of their waking hours when at home) tended to be highly correlated with $PM_{2.5}$ concentrations in the kitchen, garage, and outdoors. Living room concentrations were highly correlated with kitchen concentrations at all hours of the day for most homes. In most homes, living room $PM_{2.5}$ concentrations were also strongly correlated with both garage and outdoor concentrations in the morning. Correlation coefficients became lower throughout day before rising again in the evening and overnight. This trend was also observed with TVOC concentrations in some homes (Figure 17), although not in as many homes as the $PM_{2.5}$ correlation trend.

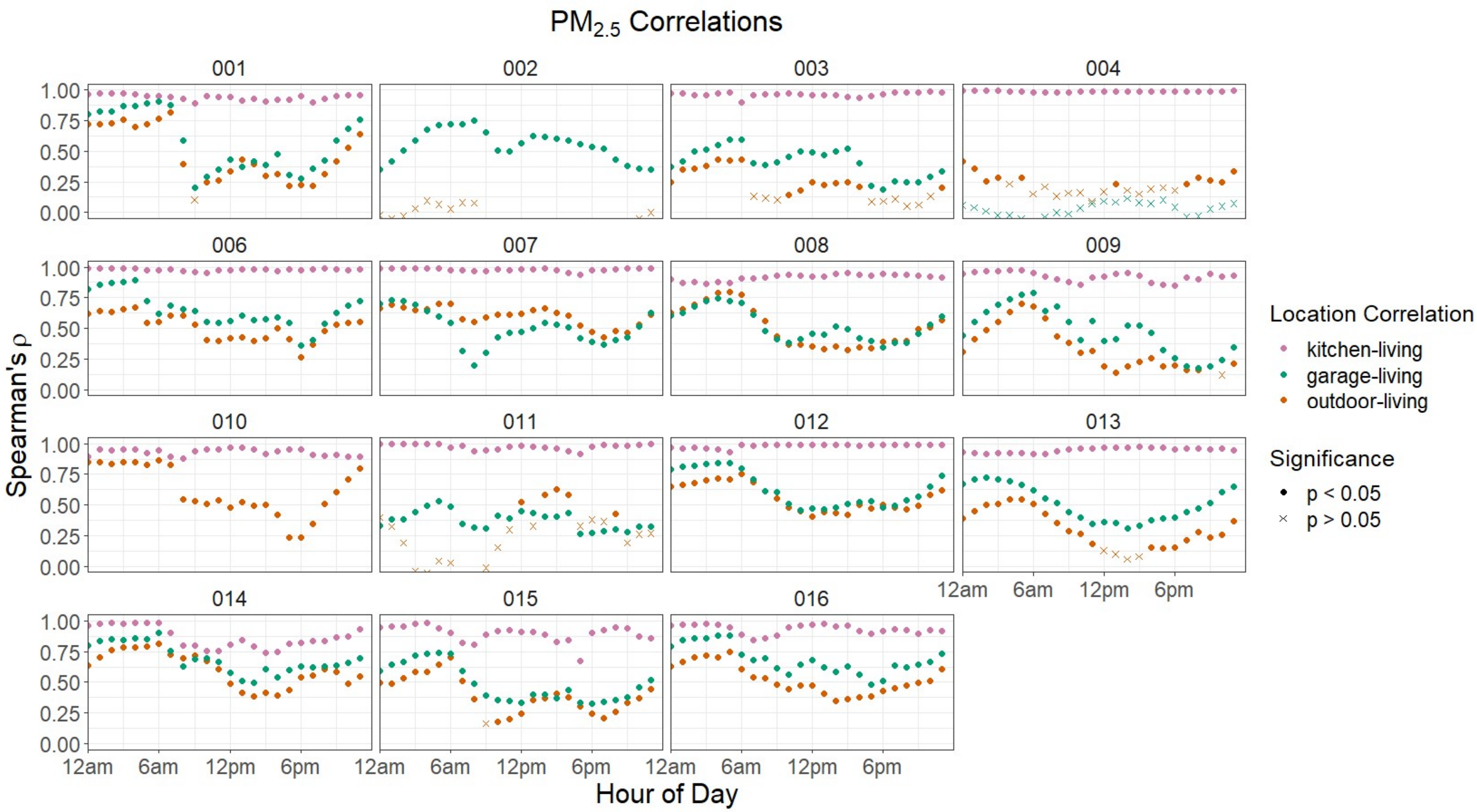


Figure 16: Pearson correlation coefficients between $PM_{2.5}$ concentrations in living room and other locations for all homes, calculated separately for each hour of day over the entire monitoring period (six to ten months). Point implies significant correlation (p<0.05); "x" implies insignificant correlation (p>0.05). No correlations were calculated if a home had no available data for a location in the considered location pair. Y-axis was restricted to positive values, as any negative correlations between rooms were assumed to be insignificant with no feasible drivers.

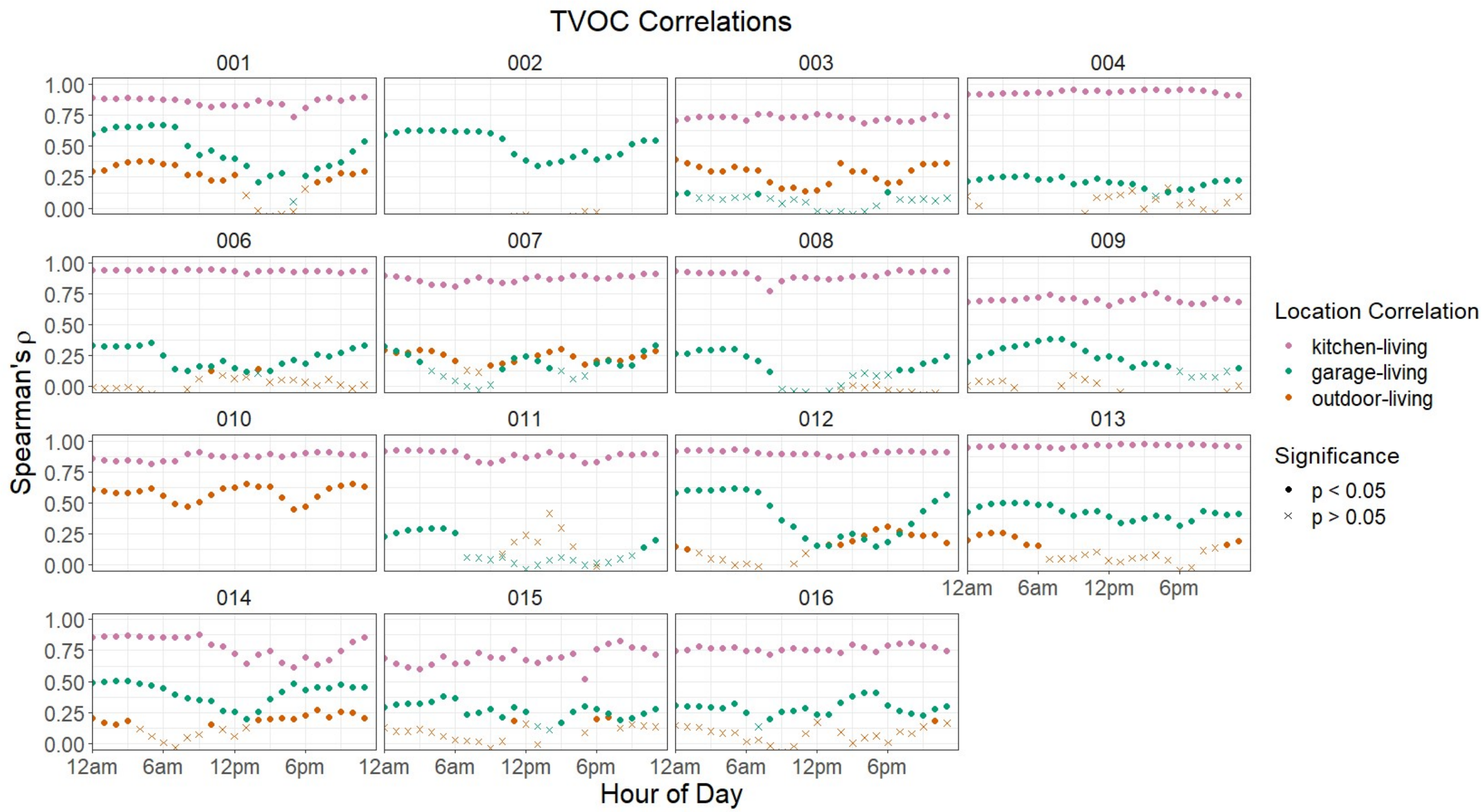


Figure 17: Pearson correlation coefficients between TVOC concentrations in living room and other locations for all homes, calculated separately for each hour of day over the entire monitoring period (six to ten months). Point implies significant correlation (p<0.05); "x" implies insignificant correlation (p>0.05). No correlations were calculated if a home had no available data for a location in the considered location pair. Y-axis was restricted to positive values, as any negative correlations between rooms were assumed to be insignificant with no feasible drivers.

As the garage, outdoors, and kitchen could be sources of indoor pollutants, the hour-of-day correlations between the garage, outdoor, and kitchen sensors for each home were explored for $PM_{2.5}$ concentrations (Figure 18) and TVOC concentrations (Figure 19). Correlations between the bedroom and garage were also included to observe possible differences with respect to the garage-kitchen results. Correlations between garage and outdoor $PM_{2.5}$ concentrations were strong in most homes throughout most hours of the day (Figure 18). Correlations between $PM_{2.5}$ values in the kitchen and the garage for many homes had diel trends similar to those observed between the living room and garage (Figure 16), where values decreased throughout the day and increased overnight. Correlations between outdoor concentrations and garage concentrations of TVOC (Figure 19) were lower than those for $PM_{2.5}$ (Figure 18). Similar to correlations between TVOC concentrations in the garage and the indoor rooms (Figure 17) a trend occurs between the garage and the bedroom TVOC concentrations in some homes, but the trend is not as significant as the trend seen with $PM_{2.5}$ between garage and indoor rooms in most homes. Correlations between the bedroom and the garage behaved similarly to correlations between the kitchen and the garage for both $PM_{2.5}$ and TVOC.

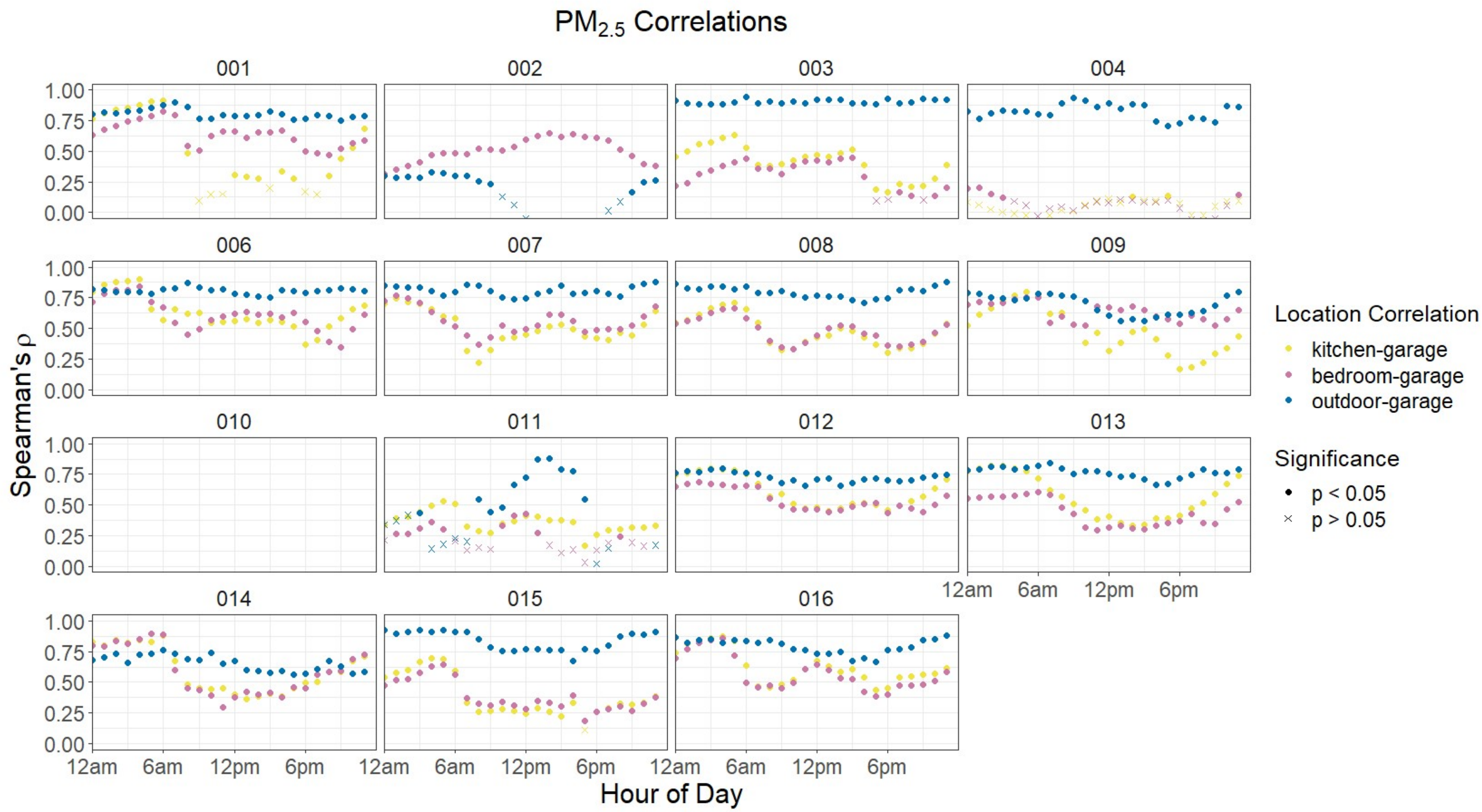


Figure 18: Pearson correlation coefficients between $PM_{2.5}$ concentrations in garage and other locations for all homes, calculated separately for each hour of day over the entire monitoring period (six to ten months). Point implies significant correlation (p<0.05); "x" implies insignificant correlation (p>0.05). No correlations were calculated if a home had no available data for a location in the considered location pair. Y-axis was restricted to positive values, as any negative correlations between rooms were assumed to be insignificant with no feasible drivers.

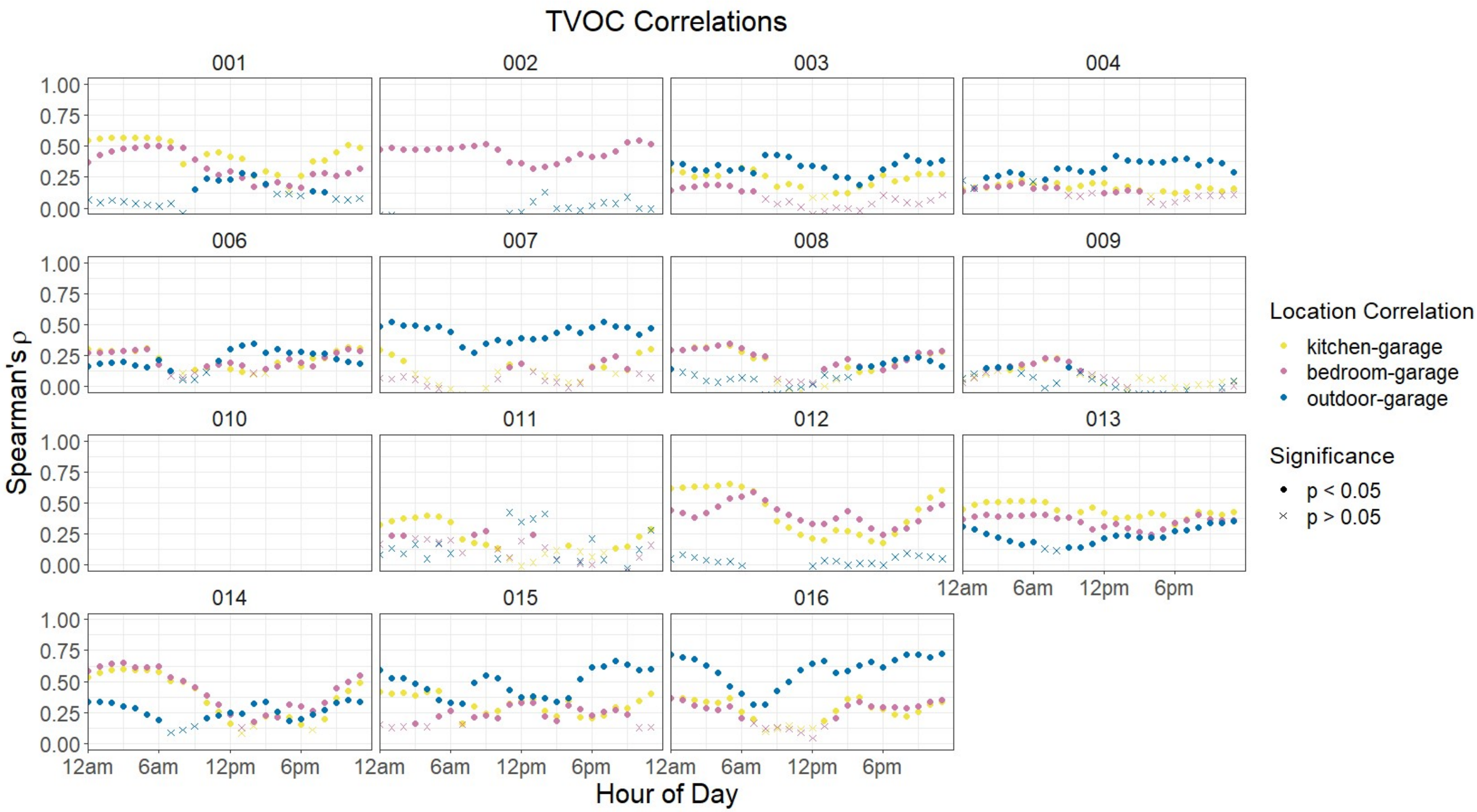


Figure 19: Pearson correlation coefficients between TVOC concentrations in garage and other locations for all homes, calculated separately for each hour of day over the entire monitoring period (six to ten months). Point implies significant correlation (p<0.05); "x" implies insignificant correlation (p>0.05). No correlations were calculated if a home had no available data for a location in the considered location pair. Y-axis was restricted to positive values, as any negative correlations between rooms were assumed to be insignificant with no feasible drivers.

5.6.4. Correlations Between IEQ Metrics

Indoor concentrations of CO_2 may sometimes be used as a proxy for occupant presence and has been used to govern ventilation in buildings (Wei et al., 2020). Pearson correlation coefficients were calculated between CO_2 concentrations and other IEQ variables that may have potential as proxies for resident presence, including TVOC, noise (measured in decibels), and light (measured in lumens) (Figure 20). TVOC concentrations tended to be moderately correlated with CO_2 concentrations for most hours of the day in most homes. Noise levels ranged from being not correlated (more often overnight) to being weakly correlated (more often during the day) with CO_2 concentrations for most homes. Noise-to-CO_2 correlations may have been present during the day because occupants likely make noise (through general activities) and generate CO_2 (through exhalation) when they are home during the day. If occupants are home at night, they would generate CO_2, but they would be less likely to make noise (especially during typical sleeping hours). Light-to-CO_2 correlations were also lower than TVOC-to-CO_2 correlations. Light-to-CO_2 correlations increased in some homes in the late evenings, and in less homes, in the mornings. Light-to-CO_2 correlations may have increased in the evening because there is typically more natural lighting present within a home during the day than overnight. Indoor light intensity during the day may be more correlated with natural outdoor light intensity than it is correlated within human activity (represented by CO_2 concentrations). Light intensity would only be expected to increase at night within a home when an occupant turns on the lights indoors, therefore causing elevated correlations of hourly CO_2 and light levels at night.

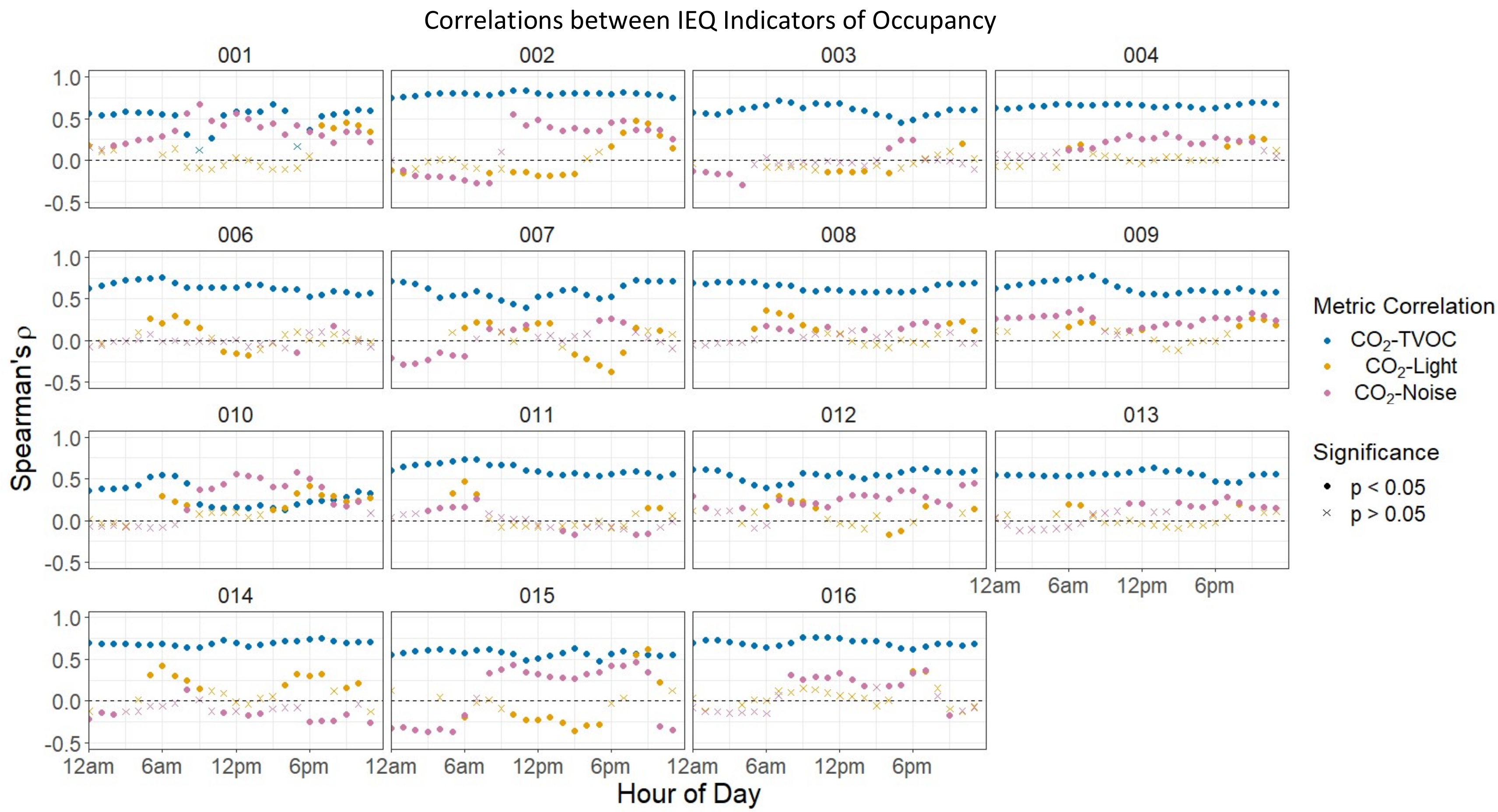


Figure 20: Pearson correlations between CO_2 concentrations and other possible metrics that could be used as resident activity proxies (TVOC and light) in the living room for each home, calculated separately for each hour of day over the entire monitoring period (six to ten months). Point implies significant correlation (p<0.05); "x" implies insignificant correlation (p>0.05).

5.6.5. Monthly Trends

Monthly average values for multiple IEQ metrics for all homes pooled together were aggregated by monitoring location to explore general monthly trends (Figure 21). Monthly average values were also shown by home (Figure 37 in Appendix). Elevated $PM_{2.5}$ levels were observed from August 2020 to October 2020 for both indoor and outdoor sensors. This was likely due to the Cameron Peak forest fire. This fire covered more than 200,000 acres at its peak, reached within 10 miles of the western border of Fort Collins, and had its most significant impact on air quality from late August to late October. Forest fires occurring in other Western states, including California, also had significant impacts on air quality in the area during these months in 2020. A second peak in $PM_{2.5}$ was observed from outdoor sensors in most homes in February and March of 2021 that was lower in magnitude compared to the fire season peaks.

TVOC indoor concentrations peaked in September and October of 2020 in many homes and then tended to decrease until January before gradually increasing through to the end of the monitoring period (May, for most homes). Garage TVOC levels were elevated relative to other monitoring locations and were more variable between homes. Outdoor TVOC concentrations were low for all homes over all months.

CO_2 concentrations in indoor rooms peaked around October 2020 and April/May 2021, and had higher averages in the bedroom than living rooms or kitchens for some homes. Lastly, even though most homes in the study reported having and using air conditioning in the summer months, average indoor temperatures were higher in July, August, and September than in other months. Low temperatures in one bedroom and one living room may be indicative of periods of vacancy for those homes or uneven heating of rooms (either by choice or necessity) (Figure 37 in Appendix).

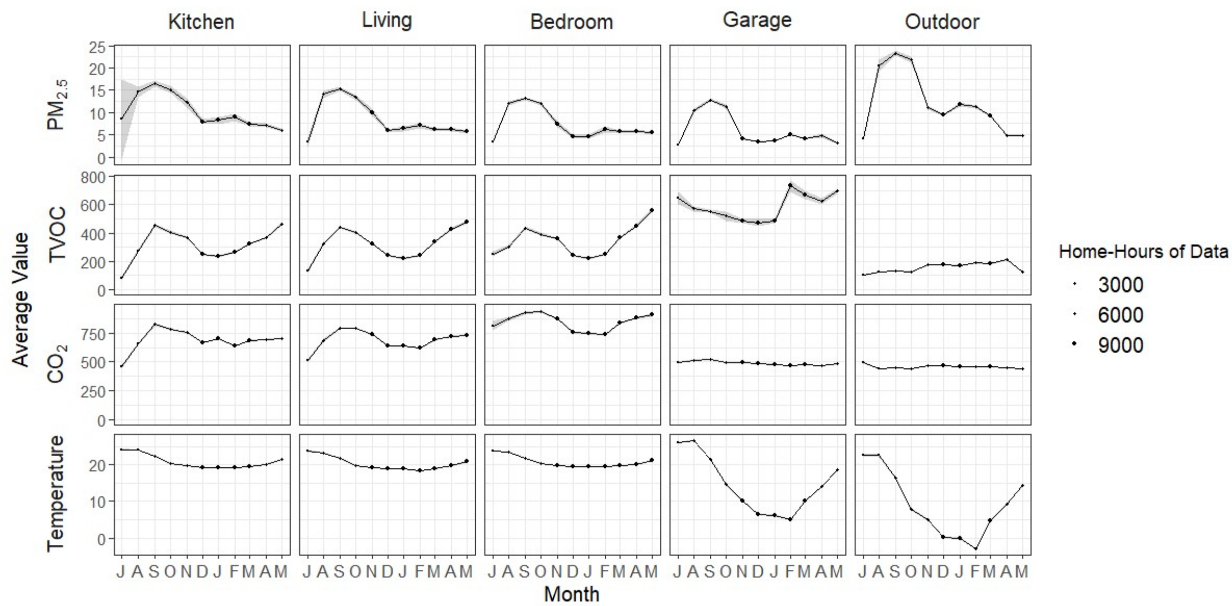


Figure 21: Monthly average values for metrics recorded over the entire monitoring period ranging from July 2020 (J) to May 2021 (M), for all monitoring locations, with data from homes pooled together. Lines represent the arithmetic monthly mean. Transparent ribbon represents the 95% confidence interval around the mean calculated from average hourly values. Top: $PM_{2.5}$ ($\mu g/m^3$). 2nd from top: TVOC (ppb). 2nd from bottom: CO_2 (ppm). Bottom: temperature (^oC).

5.6.6. Deviation from Pooled Mean Between Rooms

Differences were explored for mean hour-of-day IEQ indicator values between rooms at the home level. We calculated the hour-of-day mean values for each room separately. We then examined the difference between each room mean and the pooled (between all rooms) mean for each hour of day (Figure 22, Figure 23, Figure 24 for $PM_{2.5}$, CO_2, and temperature, respectively; TVOC shown in Figure 38 in Appendix).

In many homes, $PM_{2.5}$ kitchen mean concentrations increased above pooled mean around 6 pm (Figure 22). Living room values in most of these homes tended to increase above the pooled mean after or around the same time; however, in some homes, the bedroom increased above the pooled mean instead of the living room around 6pm. This evening deviation in $PM_{2.5}$ concentration from the pooled mean of indoor rooms was often between 3 and 10 μg/m^3 (and greater in some homes). This deviation was notable, considering the time weighted average $PM_{2.5}$ values for the rooms in most homes were less than 13 μg/m^3. $PM_{2.5}$ deviations between rooms also occurred in some homes around 9 am, although these deviations were not always as large in magnitude as the evening deviations. Most homes tended had small differences between room-specific means and pooled means of indoor rooms for CO_2 throughout most daylight hours (Figure 23). At night in most homes, mean bedroom CO_2 levels rose dramatically compared to the other two rooms. In some homes, room-specific means of temperature were, on average, nearly equal to pooled means of indoor rooms throughout all hours of the day (Figure 24). In other homes, however, the difference between room-specific means and pooled means of indoor rooms approached or exceeded ±2 °C (±3.6 °F) for many hours of the day. In many (but not all) of these homes, the bedroom temperature was higher than the pooled mean temperature over all hours of the day.

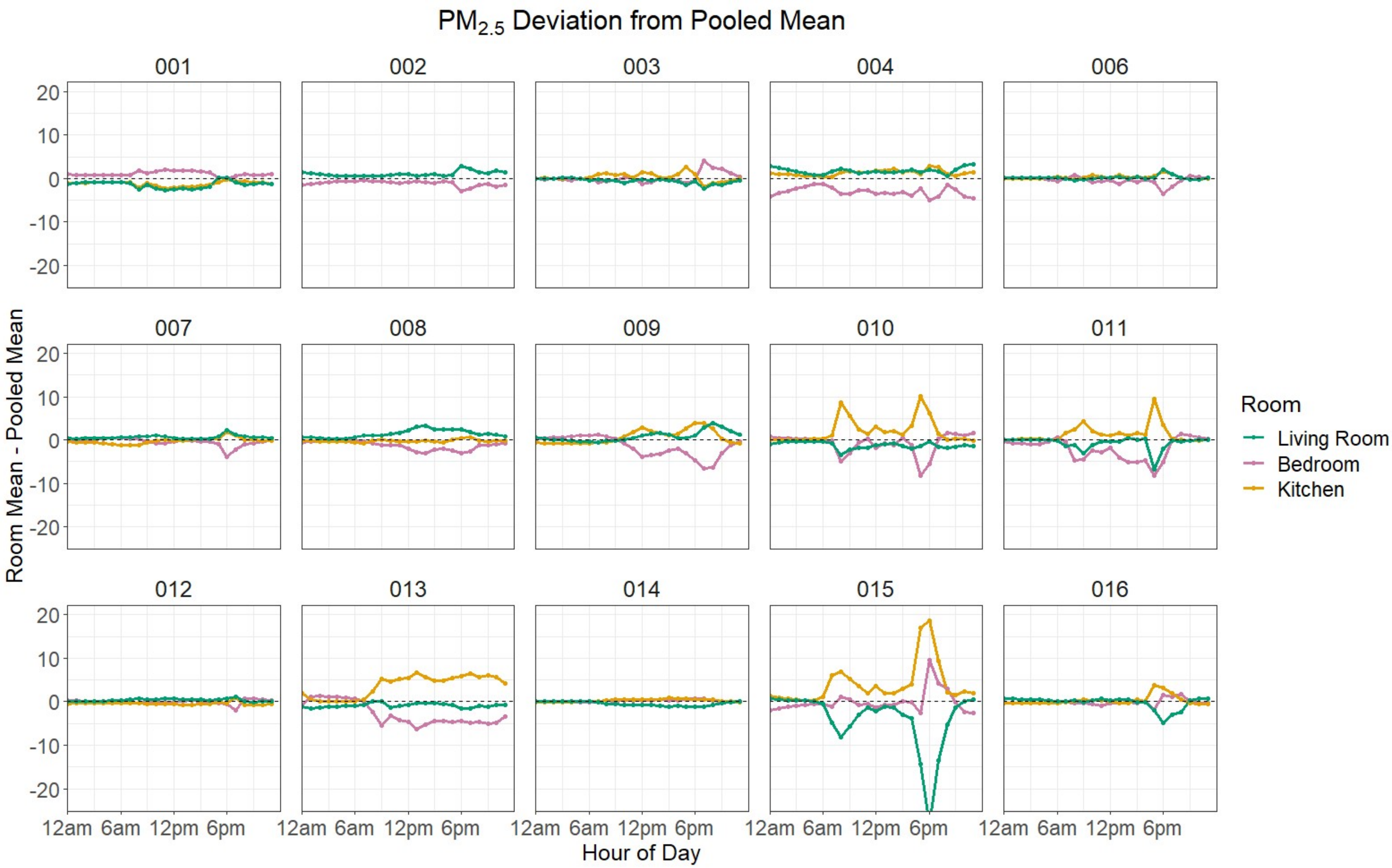


Figure 22: Pooled (between kitchen, living room, and bedroom) hour of day $PM_{2.5}$ arithmetic mean subtracted from the hour of day $PM_{2.5}$ arithmetic mean for each room at each hour of day over the entire monitoring period (six to ten months), plotted separately for all homes. Units in µg/m³. Confidence intervals were omitted to increase clarity.

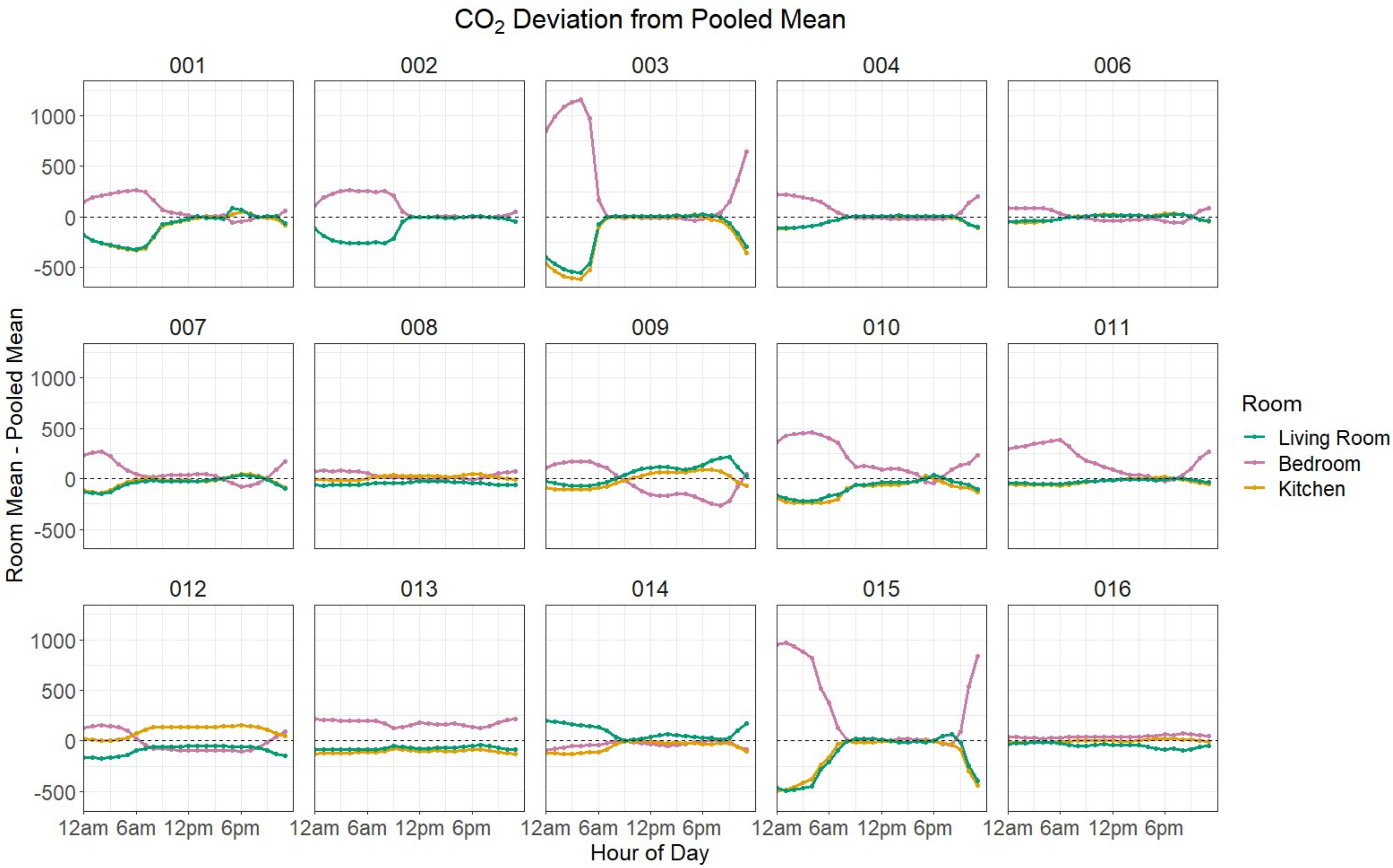


Figure 23: Pooled (between kitchen, living room, and bedroom) hour of day CO_2 arithmetic mean subtracted from the hour of day CO_2 arithmetic mean for each room at each hour of day over the entire monitoring period (six to ten months), plotted separately for all homes. Units in ppm. Confidence intervals were omitted to increase clarity.

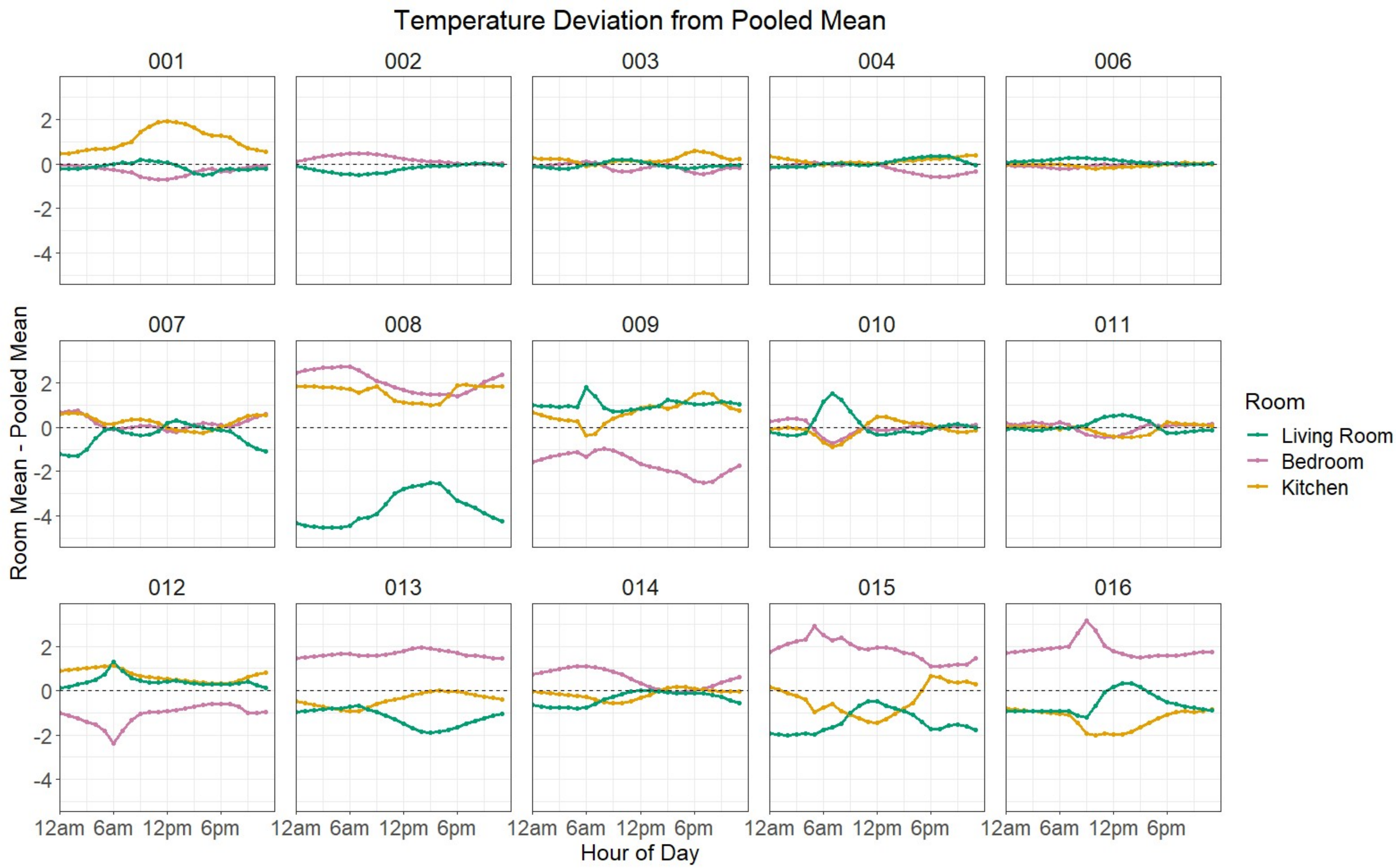


Figure 24: Pooled (between kitchen, living room, and bedroom) hour of day temperature arithmetic mean subtracted from the hour of day temperature arithmetic mean for each room at each hour of day over the entire monitoring period (six to ten months), plotted separately for all homes. Units in ºC. Confidence intervals were omitted to increase clarity.

5.7. Temporal Representativeness

Diel plots demonstrated how IEQ metrics behaved, on average, over the entire – i.e., long-term – monitoring period. Time-structured temporal representativeness was evaluated to explore how well samples taken from short-term periods (one to 28 days) represented this average diel behavior.

Temporal representativeness was determined for $PM_{2.5}$, TVOC, and CO_2 samples of lengths spanning from one to 28 days, with reference to the entire monitoring period. Samples were disaggregated by sample length, season, home, and monitor location, and the relative entropy value was determined for each sample with reference to its entire monitoring period. Within each sample length (one to 28 days) for each IEQ indicator, outlier samples were identified if they had a relative entropy value outside the range of the median relative entropy value ± 3×IQR of the relative entropy values. Outlier samples were omitted from further temporal representativeness analysis. This criterion resulted in the omission of 2.5% of $PM_{2.5}$ data samples, 5.7% of TVOC samples, and 6.0% of CO_2 samples. For each IEQ indicator, the maximum relative entropy value from the remaining samples (D_{max} in Equation 5, in Section 4.3.8.3) was used to scale values. Each season defined for each home in representative analysis had a start and end date to allow for the definition of continuous time periods (unlike the season days in the diel analysis that were defined on a day-by-day basis) (methodology described in Section 4.3.6).

To illustrate how a measure of temporal representativeness as we apply here can provide insight, we considered sample lengths of three days (i.e., 3-day), one week (i.e., 7-day), and two weeks (i.e., 14-day) for $PM_{2.5}$ (Figure 25a), TVOC (Figure 25b), and CO_2 (Figure 25c) in the living room of a single home (Home 9) for a single season (heating). The same illustration was also provided including one day (i.e., 1-day) sample lengths (Figure 39 in Appendix). We compared the temporal structure of these samples of varying lengths to the temporal structure of the entire monitoring period for each of the IEQ indicators. For $PM_{2.5}$ concentrations, the structure of all samples (3-day, 7-day, and 14-day) and the long-term period each yielded peaks two peaks: one at midday (between 10 am and 1 pm) and one in the evening (between 6 pm and 8 pm). The 3-day $PM_{2.5}$ sample midday peak occurred one hour earlier than that of the long-term period (12 pm instead of 1 pm), and the evening peak of the 3-day sample occurred two hours earlier than that of the long-term period (6 pm instead of 8 pm). Both peaks in the 3-day sample structure yielded higher concentrations than the peaks in the long-term period structure. The 7-day $PM_{2.5}$ sample was more representative than the 3-day sample, even though the evening peak of the 7-day sample was lower in magnitude and still offset from that of the long-term sample (7 pm instead of 8 pm). The 7-day sample structure had a peak at 1 pm that was more similar in magnitude to the 1 pm peak of the long-term period when compared to the 3-day sample structure 12 pm peak. The temporal structure of the 14-day $PM_{2.5}$ sample was the most representative of all three considered samples. The timing and magnitudes of the 14-day sample midday (10 am) and evening (7 pm) peaks were not always more similar to those of the long-term period when compared to the peaks of the 3-day and 7-day samples. However, the 14-day sample was the only sample out of the three sample lengths for which the value of the evening $PM_{2.5}$ peak was higher than the midday peak. The evening $PM_{2.5}$ peak of the entire monitoring period was also higher than its midday peak, which is likely why the 14-day sample structure was most representative. Similar conclusions were drawn from the TVOC and CO_2 samples when their temporal structures were compared to that of the entire monitoring period. All the TVOC sample lengths and the long-term period yielded similar timing for late evening peaks (between 8 pm and 9 pm), with the 3-day, 7-day, and 14-day TVOC samples increasing in representativeness from 3-day to 7-day to 14-day. All the CO_2 sample lengths and the long-term period yielded similar timing for midday peaks (between 12 pm and 2 pm) and late evening peaks (between 8 pm and 9pm), with the 3-day, 7-day, and 14-day CO_2 samples increasing in representativeness from 3-day to 7-day to 14-day.

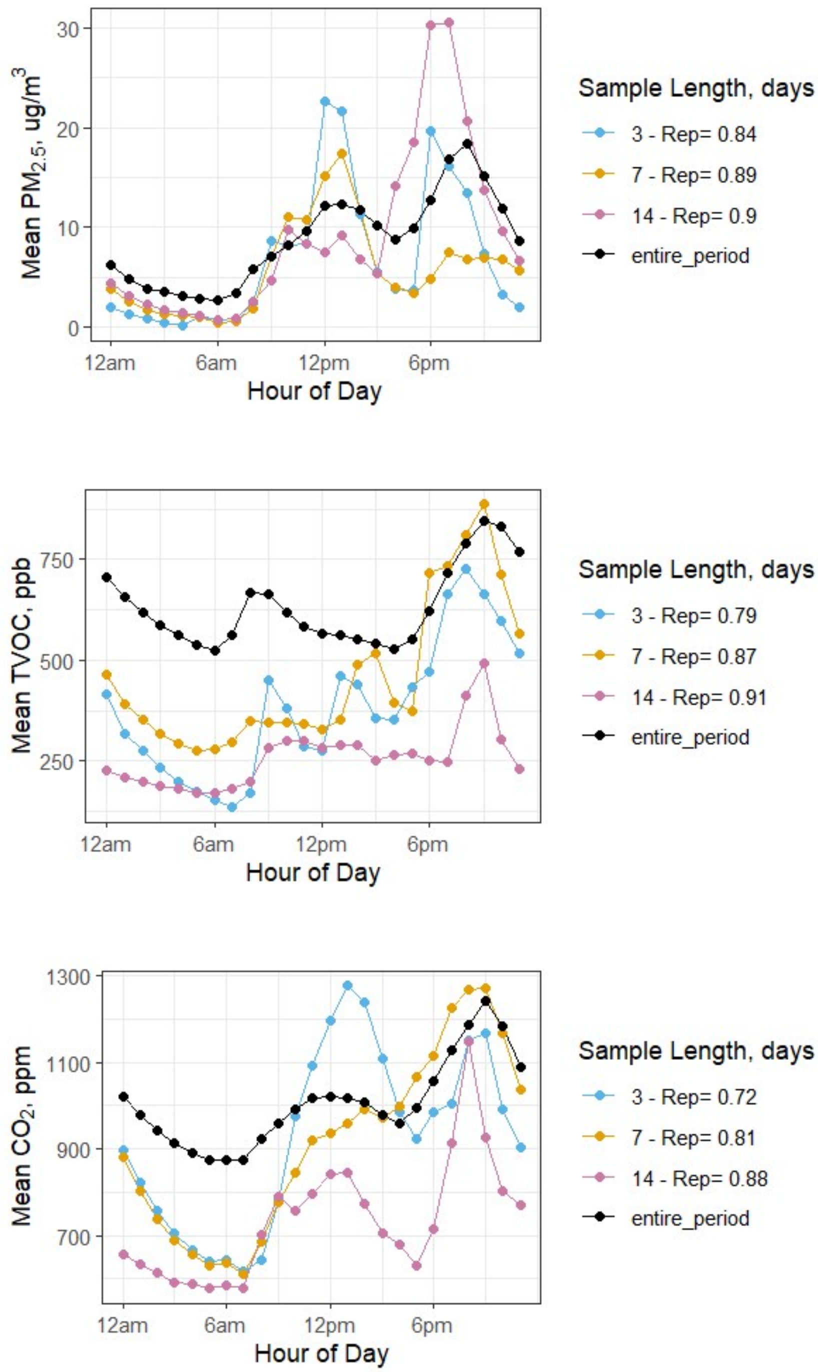


Figure 25: Hour-of-day arithmetic mean values calculated from samples of multiple lengths compared to those calculated from the entire monitoring period from the same condition for $PM_{2.5}$ (a, top), TVOCs (b, middle), and CO_2 (c, bottom). The considered condition was Home 9 living room in the heating season. Time-structured temporal representativeness (Rep) values for each sample are noted in the legend. Confidence intervals around mean values were omitted for clarity.

We analyzed how time structured representativeness of all indoor samples varies with sample length for each IEQ indicator (Figure 26). Only data from homes for which cooling seasons were identified (Homes 1, 2, 7, 9, and 12) were included in this analysis. Indoor samples of lengths ranging from one day to 28 days were extracted from each sensor's long-term (six- to ten- month) monitoring period. We extracted these samples from each season using the method described previously (Section 4.3.8.3). Only samples with lengths of one day (i.e., 1-day), three days (i.e., 3-day), one week (i.e., 7-day), and two weeks (i.e., 14-day) were included in this analysis. The representativeness for each individual sample, with respect to the long-term period from which it was extracted, was calculated with the method described previously (Section 4.3.8.3). Representativeness values of samples from all rooms in all homes were pooled, then disaggregated by season. The distribution of representativeness values for each sample length was plotted separately within each season. We selected two representativeness values (0.8 and 0.9) as possible thresholds to define when a sample was considered representative. Maciejewska & Szczurek (2015) used 0.9 as a temporal representativeness threshold in their analysis of representativeness. We selected 0.8 as another threshold option upon visual inspection of the relationship between hour-of-day averages and representativeness in the heating season (Figure 25 above, and Figure 39 in Appendix). Select samples had a representativeness value lower than 0.8: 3-day CO_2, 3-day TVOC samples, and 1-day samples of each IEQ indicator (1-day samples shown in Figure 39 of Appendix). Samples with values below 0.8 yielded average hour-of-day peaks that were notably less similar in proportion to the average peaks of the long-term period, compared to samples with representativeness greater than 0.8. For each season and sample length, the percentages of samples that exceeded each representativeness threshold was calculated (Figure 26).

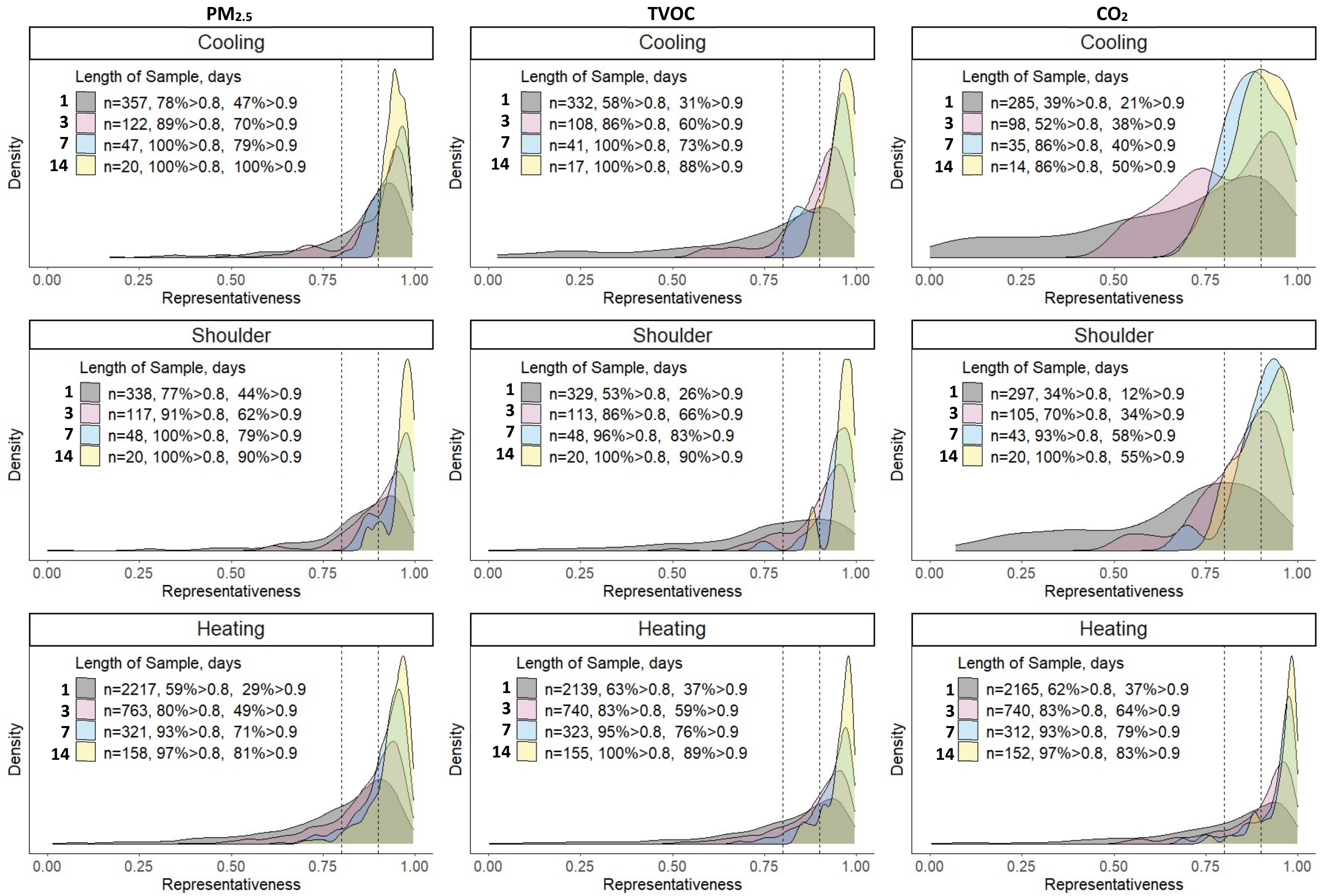


Figure 26: Density plots of temporal representativeness of $PM_{2.5}$ samples (left column), TVOC samples (center column), and CO_2 samples (right column) disaggregated by season. For each season, samples from homes and rooms were pooled then disaggregated by sample length. Only Group 3 homes included to allow between comparison of data from all three behavior-defined seasons.

The distributions of time-structured representativeness skewed closer to 1 as sample length increased, meaning longer sampling periods generally had higher proportions of representative samples than shorter sampling periods. Based on a representativeness threshold of 0.8, the proportion of 1-day $PM_{2.5}$ samples considered to be representative ranged from 59% to 78%, varying by season. This range of proportion in representative samples was higher for each increase in sample length: 80% to 91% of 3-day samples, 93% to 100% of 7-day samples, and 97% to 100% of 14-day samples. The proportion of samples that were considered representative increased with higher sample length for TVOC and CO_2 as well.

In the cooling season, based on a representativeness threshold of 0.8, the proportion of 1-day $PM_{2.5}$ samples considered representative (78%) was greater than that of 1-day TVOC samples (58%), which was greater than that of CO_2 samples (39%). This pattern ($PM_{2.5}$ > TVOC > CO_2) was true when matching by any sample length in the cooling season, save for 7-day and 14-day samples (for which, proportions of representative $PM_{2.5}$ and TVOC samples were both 100%). The proportion of representative samples pattern ($PM_{2.5}$ > TVOC > CO_2) was also present in the shoulder season when samples were matched by sample length (considering a representativeness threshold of 0.8). The proportion of representative $PM_{2.5}$ samples in the heating season (ranging from 59% to 97% between sample length) was lower compared to samples in other seasons, when matched by sample length. In contrast, the proportion of representative TVOC samples stayed relatively consistent across seasons when matched by sample length. The proportion of representative CO_2 samples was lower in the cooling season compared to samples from other seasons when matched by sample length.

Using a representativeness threshold value of 0.9 for samples to be considered representative, the same patterns described above were generally observed: representativeness increased with sample length, representativeness of samples for $PM_{2.5}$ > TVOC > CO_2 in heating and cooling season, less representative $PM_{2.5}$ samples in the heating season were less representative compared to other seasons, and CO_2

samples in the cooling season were less representative compared to other seasons. The proportion of samples considered representative for each IEQ indicator across all seasons and sample lengths decreased, as was expected when using a stricter criterion. Considering a 0.9 representativeness threshold, less than 50% of all 1-day samples were considered representative (for all IEQ indicators from any season), and less than 100% of 14-day samples were considered representative for each IEQ indicator in each season, save for $PM_{2.5}$ samples in the cooling season.

Upon visual inspection of example samples (Figure 25 above, and Figure 39 in Appendix), the IEQ samples with representativeness values below 0.8 yielded average hour-of-day peaks that were notably different in timing and magnitude with respect to long-term monitoring period average hour-of-day peaks. As representativeness of samples increased above 0.8, average hour-of-day structures varied less in similarity to that of the long-term monitoring period. This threshold of 0.8 was chosen as the preferred temporal representativeness threshold based on the results of this study. However, the implications of temporal representativeness values assigned to each sample using the methods from the current study should be further explored before a more certain threshold is established.

5.8. Spatial Representativeness and Specificity

The method used to measure spatial representativeness and specificity required each considered dataset to meet the assumption of normality. QQ plots were created for each home-room-season condition for each IEQ indicator considered (Figure 40, Figure 41, and Figure 42 in Appendix). Most IEQ parameter samples were log-normally distributed, and transforming the distributions with the natural log yielded approximately normal distributions. The raw data for each IEQ parameter were fitted to log-normal, gamma, and Weibull distributions to test if other distributions fit better than log-normal. Log-normal distributions yielded the best fit under most conditions for most of the IEQ indicators considered according to the Akaike information criterion (AIC) (Table 12 in Appendix). Note, in the rest of this study, any references to the spatial representativeness or specificity of an IEQ indicator were calculated from the natural log of the indicator's values.

Using methods described previously (Section 4.3.8.4), measures of spatial representativeness and specificity were calculated for $PM_{2.5}$, TVOC, and CO_2 concentrations recorded by each indoor sensor over each season in each home. Spatial representativeness and specificity values were compared between rooms, disaggregated by season (Figure 27). Bedroom measurements had the lowest median representativeness of all three rooms for all three IEQ indicators. Median living room TVOC measurements were more representative than median TVOC measurements in kitchens, while the spatial representativeness of living rooms and kitchens did not differ for $PM_{2.5}$ and CO_2 measurements. The median spatial specificity for bedroom was higher than for the kitchen or living room for $PM_{2.5}$, TVOC, and CO_2 measurements; living room and kitchen specificity values had similar low magnitudes. No discernible difference between season for representativeness nor specificity values was observed.

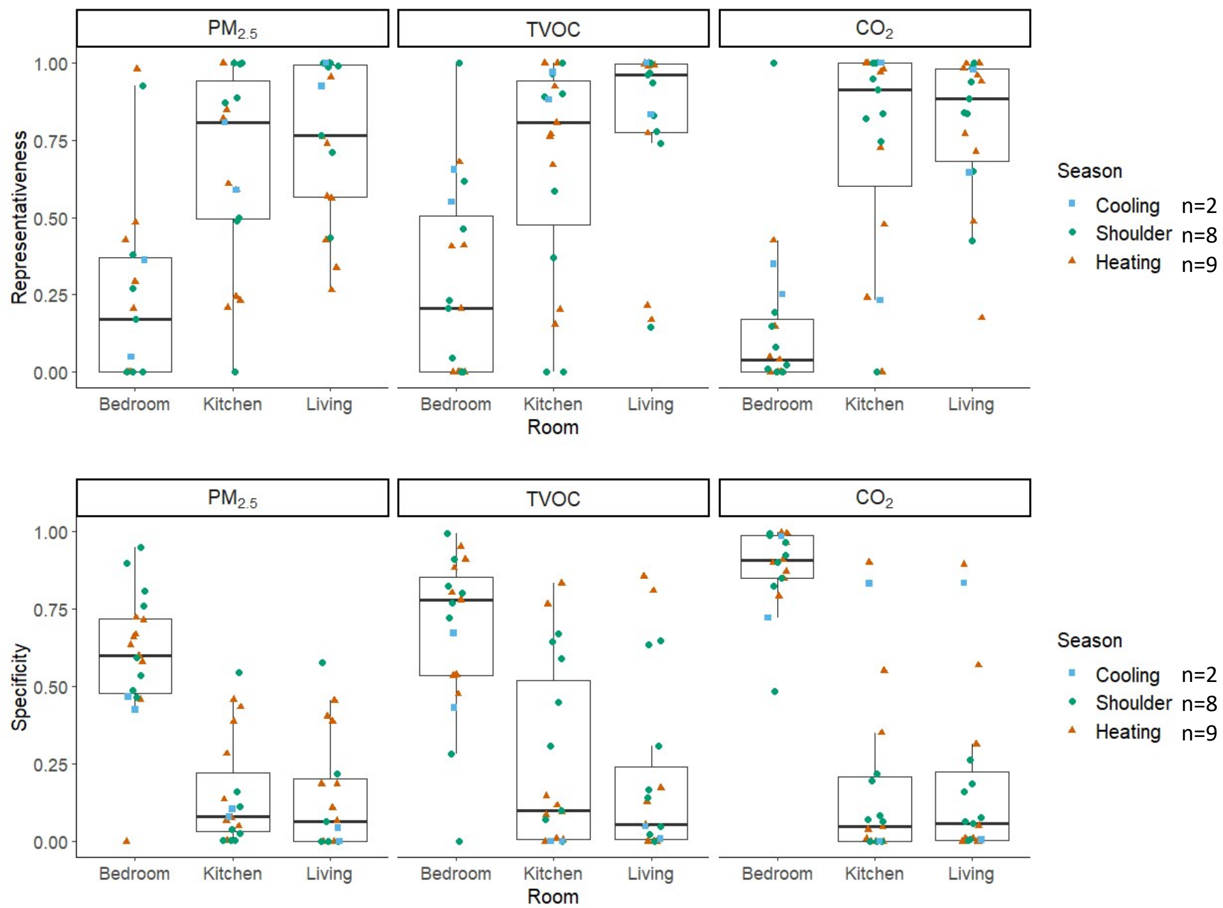


Figure 27: For IEQ indicators of $PM_{2.5}$, TVOC, and CO_2: spatial representativeness (top) and spatial specificity (bottom) values calculated for all indoor room sensors over each entire behavior-defined season within each home, disaggregated by room and colored by season, homes pooled. For each season, only sensors from homes that had data from all three rooms were included. n in legend shows number of homes included for each season for each distribution.

Absolute values of spatial representativeness and specificity (Figure 27) do not have interpretative value. However, ranking the three indoor IEQ monitoring locations from least to most representative/specific within each home (Figure 28 and Figure 29) provides insight into the representativeness and specificity of each monitoring location relative to one another within the same home. For $PM_{2.5}$ and TVOC concentrations in most homes, bedroom sensor provided data that was least representative of the entire home's data but the most specific. $PM_{2.5}$ concentrations measured in the kitchen in most homes were more specific than $PM_{2.5}$ concentrations measured in corresponding living rooms. No notable seasonal differences were observed for specificity or representativeness rank of any IEQ indicator.

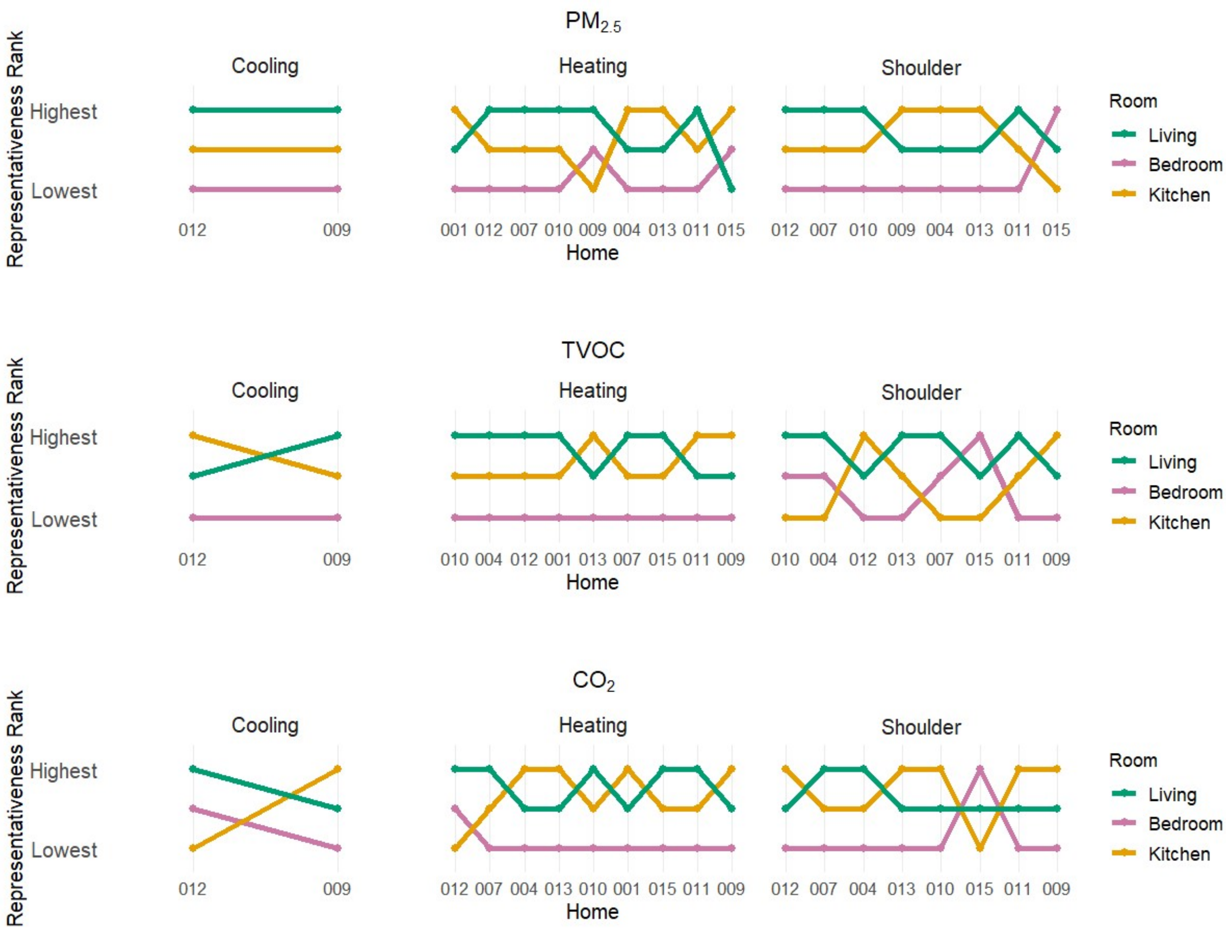


Figure 28: Within-home ranks of spatial representativeness for each room calculated for all indoor sensors over each entire behavior-defined season within each home, disaggregated by season for PM2.5 (top), TVOC (middle), and CO_2 (bottom). For each home, the top room had the highest representativeness value of all three rooms, and the bottom room had the lowest representativeness value of all three rooms. For each season, axes are ordered by homes with lowest time-weighted average on the left, to homes with highest time-weighted average on the right for the considered pollutant, and only sensors from homes that had data from all three rooms were included.

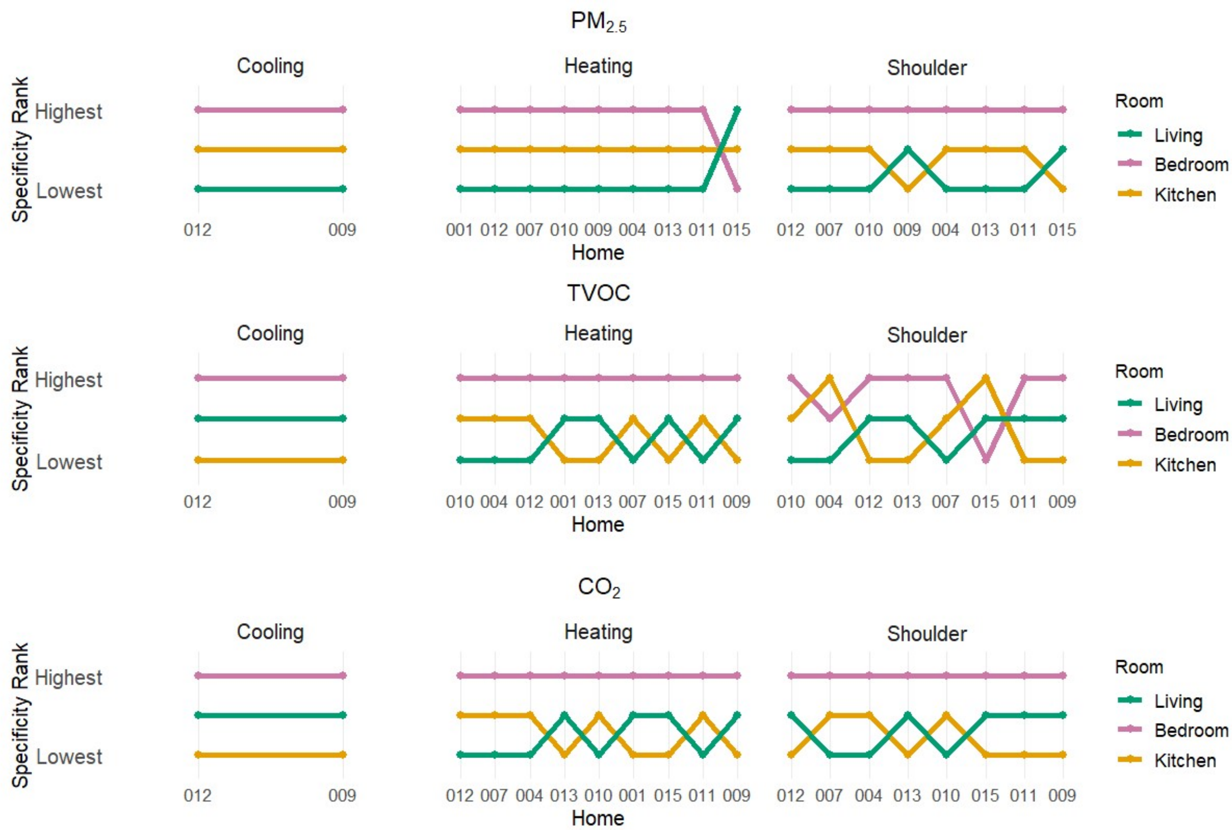


Figure 29: Within-home ranks of spatial specificity for each room calculated for all indoor sensors over each entire behavior-defined season within each home, disaggregated by season for $PM_{2.5}$ (top), TVOC (middle), and CO_2 (bottom). For each home, the top room had the highest specificity value of all three rooms, and the bottom room had the lowest specificity value of all three rooms. For each season, axes are ordered by homes with lowest time-weighted average on the left, to homes with highest time-weighted average on the right for the considered pollutant, and only sensors from homes that had data from all three rooms were included.

6. Discussion

6.1. Temporal Trends

6.1.1. Season Classifications

Occupant behavior has been shown in IEQ literature to impact IEQ indicators in homes. Heating device analysis performed in the present study (Figure 4) suggested using traditional season start and end dates (e.g., June 21st or March 21st) to define seasons for all occupants in Colorado-based residential IEQ studies may not properly account for variables related to occupant behaviors. Some homes were shown to lack continuous periods of heating and cooling (i.e., heating days were not always preceded and followed by additional heating days). We also observed between-home variability in occupant space-heating patterns. Cooling seasons, (i.e., continuous periods when residents typically use air condition) were only identified in five of the 13 analyzed homes over the study. When continuous heating or shoulder seasons were observed, the start date varied between homes. In their study developing a statistical model of heating prediction in Dutch dwellings, Macjen et al. (2020) noted occupant perception of temperature and occupant perception of air humidity affected heating. The present study noted the date of the shoulder-to-heating season transition varied more between home than that of the heating-to-shoulder transition. This variability may be related to changes in behaviors noted by Huchuk et al. (2018) and other studies: past exposures and experiences, which differ between participants, may impact perceived thermal comfort (Brager & de Dear, 1998; Richard J. de Dear & Gail Schiller Brager, 1998). Residents may consider themselves warmer in the winter than in the summer, even at the same indoor temperature (Oseland, 1994).

6.1.2. Diel Trends

On average, $PM_{2.5}$, TVOC, and CO_2 time series exhibited similar diel trends between homes and seasons in kitchens and living rooms (Figure 15). Concentrations tended to increase around typical waking hours (between 6 am and 9 am), stabilize throughout the day, and peak in the evening when more human activity may be expected (cooking, people arriving home from work or outdoor activities, etc.). Bedroom concentrations of TVOC and CO_2 exhibited average diel trends that were notably different than the other two indoor rooms for these two indicators. Concentrations of TVOC and CO_2 in the bedroom tended to increase during late evening hours, when occupants were likely sleeping. Notably, data was collected for this study when many participants were likely working from home due to COVID-19 restrictions. Therefore, researchers may benefit from comparing these data with other high-resolution datasets collected in other similar locations during more "typical" periods (e.g., when most residents have not been encouraged to work from home).

In their study on airborne particulate matter and bacteria, Clements et al. (2018) recorded 115, 24-hour $PM_{2.5}$ samples within fifteen Colorado homes. This resulted in a mean ± standard deviation of 8.1 $\mu g/m^3$ ± 8.1 $\mu g/m^3$ across all 24-hour samples. The results of the current study over the entire monitoring period were comparable (Table 9). Time-weighted six- to ten- month average $PM_{2.5}$ concentrations calculated for each home in the current study yielded a median value of 7.7 $\mu g/m^3$, ranging from 2.4 $\mu g/m^3$ to 16.3 $\mu g/m^3$ between homes. In a study in the Denver, Colorado area, Militello-Hourigan & Miller (2018) determined three- to five-day background (non-cooking event) average $PM_{2.5}$ concentrations for living spaces in ten homes one public library (nine of the homes were "tightly constructed" – i.e., had low air exchange rates). These authors recorded a median time weighted three- to five-day average value of 12.5 $\mu g/m^3$, with a range of 2.3 to 32.6 $\mu g/m^3$ between homes. Militello-Hourigan & Miller's results were comparable to the results presented in the current study, though our concentrations were generally lower.

When pooling all homes (Figure 14), we observed elevated hourly average $PM_{2.5}$ concentrations in mornings (around 9 am) and evenings (around 6 pm) in all indoor rooms, suggesting cooking-related sources of $PM_{2.5}$ may have been be dominant. It is well-studied that cooking is a source of particulate matter. The cooking of meat was determined as a source of 6.2% of outdoor $PM_{2.5}$ carbon on average in an area of Denver, Colorado (Watson et al., 1998). Studies conducted in Colorado (Clements et al., 2018; Escobedo et al., 2014; Militello-Hourigan & Miller, 2018) and globally (Abdullahi et al., 2013; Lai et al., 2010; Massey et al., 2012; Tan et al., 2013; Wan et al., 2011) have identified cooking as one of the most significant contributors to indoor PM generation. Many studies have also determined outdoor traffic to be a contributor to indoor particulate matter (Urso et al., 2015). The timing of average hour-of-day $PM_{2.5}$ peaks in the current study were consistent with expected commuting times; however, neither the outdoor sensors nor the garage sensors observed hour-of-day $PM_{2.5}$ peaks at times similar to those of the indoor sensors. This difference in elevated outdoor and indoor peak timing suggests elevated indoor hour-of-day $PM_{2.5}$ concentrations were due to indoor sources. The correlation in hourly $PM_{2.5}$ measurements recorded between pairs of indoor rooms (stronger than between outdoor and indoor pairs) also suggests that indoor sources of $PM_{2.5}$ had a more significant impact than outdoor sources. This would be consistent the results observed in a study of living areas in ten Colorado homes and one library with low air exchange rates (Militello-Hourigan & Miller, 2018). Excluding concentrations recorded during cooking periods, Militello-Hourigan & Miller observed low concentrations of indoor $PM_{2.5}$ for all homes, relative to outdoor concentrations. The authors attributed this to lack of penetration of outdoor $PM_{2.5}$ into the homes.

Studies conducted on IEQ parameters in the residential setting have included attached or built-in garage as a variable in indoor $PM_{2.5}$ regression analyses (Héroux et al., 2010; Urso et al., 2015). However, the extent to which $PM_{2.5}$ is transported between the garage and the indoors is a focus that is lacking in literature. For most homes in the current study, hourly correlation values between garage and living room $PM_{2.5}$ concentrations tended to decrease throughout the day and increase overnight (Figure 16). One interpretation of this result is transport of $PM_{2.5}$ between the garage and the home became more significant as nighttime progressed. The early morning decrease in garage-to-indoor correlations may have been due to depressurization of the garage as the door is opened. However, this trend in hourly $PM_{2.5}$ garage-to-indoor correlations was observed in homes with both attached and detached garages. An alternative interpretation for this trend in hourly correlations may be the daytime generation of particles indoors did not impact changes in outdoor or garage $PM_{2.5}$ concentrations. For example, indoor $PM_{2.5}$ concentrations were impacted by cooking and other occupant-related activities when participants were awake. Once participants went to sleep, the indoor hourly concentrations approached equilibrium with garage and outdoor concentrations. This could have caused elevated indoor-to-garage and indoor-to-outdoor nighttime correlations compared to daytime correlations. The IEQ data from the current study, or IEQ data with similar temporal resolution, could be further analyzed to explore the mechanisms through which airborne particles are (or are not) transported between living spaces and garages/outdoors. Energy use or behavioral data paired with such IEQ data would help to explore these mechanisms.

In their study of living areas in ten Colorado homes and one library with low air exchange rates, Militello-Hourigan & Miller's (2018) results yielded three- to five- day time-weighted average TVOC values concentrations ranging from 139 ppb to 426 ppb. The median household three- to five-day time-weighted average TVOC concentration was 289 ppb, although this was excluding the one conventionally built home. Time-weighted six- to ten- month average TVOC concentrations calculated for each home in the current study yielded comparable (but higher) concentrations (median value of 318 ppb, ranging from 176 ppb to 565 ppb between homes) (Table 9). When pooling all homes, we observed elevated hourly average TVOC concentrations in evenings (peaking around 7 pm) in kitchens and living rooms and later at night (peaking around 12 am) in bedrooms (Figure 14). The evening peak in kitchen and living room TVOC concentrations may have been influenced by TVOC concentrations in garages that peaked in evenings due to increasing garage temperatures throughout the day. As temperature increases, so do the volatilities of VOCs, thus increasing airborne VOC concentrations (Clarisse et al., 2003; Vardoulakis et al., 2020). Low- to moderate- correlations observed between hourly garage concentrations and living room concentrations of TVOC (Figure 16) suggested interaction between TVOC concentrations in the two locations. Insignificant- to low- correlations between hourly TVOC concentrations in the living room and outdoors were observed (Figure 16), as were low average outdoor TVOC concentrations (Figure 14). These two observations suggest indoor-to-outdoor TVOC interactions were not as significant as garage-to-indoor interactions. Past studies have found significant infiltration of air from attached garages into homes, resulting in rising concentrations of VOCs that originate from fuels and other solvents stored in the garage (Batterman et al., 2007; Dodson et al., 2008; Fugler et al., 2002; Mallach et al., 2017). This infiltration of VOCs from the garage can be due to a negative pressure differential between the garage and home. Home-to-garage pressure differentials can be impacted by the operation of home ventilation systems or temperature differences between the home and garage (Graham et al., 1999; Mallach et al., 2017).

High average hour-of-day TVOC concentrations in evenings for indoor rooms (Figure 14) may have also been due to human activities. Residential studies have noted that human activities (e.g., using cleaning products and clothes washing) are sources of select VOCs (Batterman et al., 2007; Rösch et al., 2014). CO_2 is a widely accepted proxy for building occupancy and ventilation (Bekö et al., 2016; Militello-Hourigan & Miller, 2018). As such, the association of TVOC concentrations with human activities was exemplified in this study by the moderate- to strong- hourly correlations between TVOC and CO_2 concentrations in all homes (Figure 20). Cooking has also been cited as a significant source of some VOCs (Clarisse et al., 2003). This is consistent with elevated $PM_{2.5}$ and TVOC concentrations that we observed concurrently during typical evening cooking hours (Figure 14). In a study of IAQ indicators within two

Italian homes during COVID-19 lockdown, hour-of-day TVOC concentrations in kitchens and bedrooms were attributed to human activities such as cooking and waking up/going to sleep (Pietrogrande et al., 2021). The timing of the hour-of-day peaks reported in their study (between 7 am and 10 am and between 6 pm and 7 pm for kitchens and bedrooms) partially agreed with the timing of average hour-of-day TVOC concentration peaks in the current study (7 pm in kitchens and 12 am in bedrooms).

Time-weighted six- to ten- month average CO_2 concentrations calculated for each home in the current study yielded a median value of 736 ppm, ranging from 620 ppm and 1,004 ppm between homes (Table 9). These CO_2 concentrations were lower than those reported in another study of 100 Colorado homes of Mexican immigrants (Miller et al., 2009). Miller et al. recorded a 24-hour weighted average CO_2 concentration within each home, resulting in a mean of 1170 ppm and a standard deviation of 573 ppm. Bedrooms in the current study recorded average hour-of-day CO_2 values that were elevated above 1000 ppm at night (Figure 14). This result is comparable to the study of Colorado homes conducted by Militello-Hourigan et al. (2018), in which over half the sampled homes experienced CO_2 levels above 1000 ppm overnight.

6.1.3. Long-Term Variability

Elevated average levels of outdoor $PM_{2.5}$ reported from August 2020 to October 2020 (Figure 21) were likely due to the Cameron Peak fire of 2020 and other fires occurring in the Western United States during this period. Outdoor $PM_{2.5}$ concentrations in the Central Rocky Mountain Region of the United States have been shown to correlate significantly with wild fire frequency in surrounding states (Jaffe et al., 2008). Elevated outdoor $PM_{2.5}$ from fires in August through October may have caused elevated indoor concentrations over the same period in the current study. Outdoor $PM_{2.5}$ peaks in February and March have been reported in past years in Fort Collins and other Colorado Front Range communities

(Colorado Department of Public Health & Environment, 2019). A past study in Boulder, Colorado, also located in the Front Range, reported highest concentrations of outdoor $PM_{2.5}$ in summer months (Clements et al., 2018). These authors also reported elevated outdoor concentrations in the winter months compared to spring and fall concentrations. For indoor locations, Clements et al. (2018) reported the lowest $PM_{2.5}$ concentrations in winter months. Our study reported consistent findings: the average hour-of-day indoor $PM_{2.5}$ concentrations of heating days were lower than those of cooling days (Figure 15). Clements et al. (2018) attributed higher concentrations in the summer to outdoor $PM_{2.5}$ to increased natural ventilation. Our study hypothesized less natural ventilation would occur during cooling and heating days, as residents may keep doors and windows closed during these days to retain cool or warm air. With our hypothesis, elevated cooling day $PM_{2.5}$ concentrations would have been due to infiltration of fire-produced $PM_{2.5}$ through openings in building envelopes instead of through open windows. This interpretation of elevated cooling day indoor $PM_{2.5}$ concentrations is supported by the higher average indoor $PM_{2.5}$ peaks recorded in mornings compared to evenings in the cooling season (Figure 15). For most homes in which air conditioning was detected, most cooling days occurred during the period in which the Cameron Peak fire likely impacted outdoor air quality the most (between August 13, 2020 – i.e., when the fire began – and the end of October 2020 –i.e., when monthly average outdoor sensor $PM_{2.5}$ concentrations decreased substantially) (Figure 4 and Figure 21). Outdoor sensors reported a large average hour-of-day peak $PM_{2.5}$ concentration at 9 am over the August 13, 2020 through October 2020. This 9 am peak was not present in outdoor $PM_{2.5}$ average hour-of-day concentrations when excluding concentrations recorded (Figure 34 in Appendix). This morning outdoor average $PM_{2.5}$ peak during fire-impacted days could have led to elevated cooling day indoor $PM_{2.5}$ concentrations via infiltration. We did not ask participants how they responded to fire conditions (i.e., whether or not they opened windows more often to increase ventilation, or kept windows closed to avoid smoke from entering). Therefore, we cannot conclude the cause for elevated cooling day $PM_{2.5}$ concentrations.

The elevated monthly indoor averages for TVOC concentrations reported around October (Figure 21) may have been due to participants keeping windows and doors closed to avoid outdoor air from entering during the wildfire season. VOCs may originate from participant activities, or from off-gassing of in-home paints, varnished, furnishings, and other construction/renovation-related materials (Bari et al., 2015; Batterman et al., 2007; Rösch et al., 2014; Vardoulakis et al., 2020). Subsurface soil contamination is also known to impact indoor air VOC concentrations, and the magnitude of this impact has been shown to very between seasons (Holton et al., 2013). Closed windows during wildfire season could have led compounds of indoor origin to accumulate indoors if mechanical ventilation was insufficient. The current study only measured TVOC and not individual VOCs; therefore, we have no insight on specific gas-phase air pollutant composition. Also, we did not inquire about participant response to fires, so we cannot attribute elevated TVOC concentrations to closed windows. During weatherization-related and energy assessment programs, such as the Epic Homes program, energy efficiency practitioners often recommend energy efficiency upgrades (e.g., new ventilation systems or measures or air sealing between the home and garage) to homeowners. Sources and behaviors that trap VOCs indoors are important for practitioners to consider when making recommendations. Practitioners can also inform residents on behaviors that may increase TVOC concentrations, especially when indoor air exchange rate may be low (e.g., cooling seasons or fire seasons).

The variability in hour-of-day structures for IEQ indicator samples raises important questions relating to IEQ sampling periods. Will a two-day sample of $PM_{2.5}$ have the same average hour-of-day structure as the entire year, or will between-day variability cause the timing of hour-of-day peaks to differ between the sample and annual data? Do the effects of this between-day variability vary by sampling season? The measure of temporal representativeness explored in this paper aimed to answer these questions.

6.2. Spatial Variability

Average hour-of-day $PM_{2.5}$ concentrations in the living room for many homes rose concurrently with kitchen concentrations above the overall pooled mean of all rooms. Some homes recorded increased bedroom $PM_{2.5}$ concentrations following kitchen peaks, with respect to pooled indoor means (Figure 22). In their study of $PM_{2.5}$ and particles of other size fractions emitted during cooking events of gas-powered stoves in 12 Chinese homes, Wan et al. (2011) observed that living room $PM_{2.5}$ mass concentrations increased by an average of 50% during cooking events and remained above background levels an hour after completion of cooking. As noted in Urso et al. (2015), the impact of cooking activities on living rooms (and likely other rooms in the home) depends on many factors, including ventilation systems and methods and positioning of the kitchen. These structure-related factors could explain the variability between homes.

CO_2 concentrations were highest in bedrooms at nighttime (Figure 14). Similarly, Militello-Hourigan & Miller (2018) reported elevated bedroom CO_2 concentrations within homes in the Denver, Colorado area. This observation could inform sleep studies. Elevated bedroom CO_2 concentrations underscore the potential value of data collected in multiple locations in residential environments, depending on the aim of the study. Results showed, as is expected, that a single sensor would only capture this elevated CO_2 data if placed in the bedroom (not the living room or kitchen) (Figure 7 and Figure 21). Similarly, the differences in diel room-specific means and diel pooled room means of temperature (Figure 24) suggested uneven heating/cooling of rooms within some homes. It is unknown whether this uneven heating/cooling was by choice (e.g., participants preferred a warmer bedroom), necessity (e.g., participants could not afford to heat all rooms), or home structure (e.g., hot air rose to upstairs bedrooms or rooms are not heated evenly by heating system). Future work, once more homes are recruited, could include variables related both to home structure (e.g., distance between rooms, presence of stove ventilation fan) and resident behaviors (e.g., use of stove fan, air conditioning, or heating devices) in regression analyses. These analyses may help identify factors that impact the spread of pollutants within homes and the variability of thermal comfort-related indicators between rooms.

A mean value of an IEQ indicator calculated over the entire monitoring period (or season) is not the only value worth comparing between rooms. The range of IEQ indicator values, and the hour-of-day at which an indicator peaks, can vary between rooms. Variability of IEQ indicators between rooms, considering both diel time structure and magnitude, support the need for spatial representativeness and specificity measures. These spatial measures may help determine which rooms, or how many rooms, within a home should be monitored to properly characterize home conditions. Also, between-home differences in spatial variability may be important for energy efficiency practitioners. When paired with home characteristic and behavioral data, phenomena such as uneven heating or higher pollutant concentrations in certain rooms at different hours of the day may be better understood. Informed with data of proper temporal and spatial resolution, practitioners would be more equipped to recommend upgrades and behavioral changes to homeowners that would be beneficial to occupant health.

6.3. Temporal Representativeness

According to the analysis presented in this study, in the cooling and shoulder seasons, TVOC samples generally required more days than $PM_{2.5}$ to be considered representative. CO_2 samples required the most days out of the three considered IEQ indicators ($PM_{2.5}$, TVOC, and CO_2) to be considered representative in the cooling and shoulder seasons (Figure 26). To our knowledge, no studies have compared representativeness of measurements between these three IEQ indicators. Luoma & Batterman (2000) discussed representativeness while measuring the variability of IAQ indicators (i.e., particulate matter of multiple size ranges between 0.3 and 25 µm, airborne fungi/bacteria, and CO_2). The authors posited that spaces with lower air exchange rates (thus higher autocorrelation factor values between subsequent measurements) would result in greater between-sample variability. The authors infer that when an IAQ pollutant source is introduced to a room with a low air exchange rate, pollutant sample concentrations will increase substantially from the sample mean and remain above the mean for subsequent measurements. The authors state that when the same pollutant source is introduced to a room with a high air exchange rate, sample concentrations will fluctuate more rapidly, but sample concentrations will not change substantially from the sample mean. Luoma & Batterman (2000) suggested higher between-sample variability from spaces with low air exchange (associated with higher autocorrelation factor values) would lead to less representative samples. Our study did not include air exchange measurements in our analysis, but autocorrelation was analyzed for each IEQ indicator. Daily autocorrelation factor values were lower, on average, for $PM_{2.5}$ samples compared to TVOC and CO_2 samples (Figure 9). Coefficient of variation for hourly $PM_{2.5}$ samples pooled across all homes (3.27) was higher when compared to that of TVOC samples (1.11) and CO_2 samples (0.49) (Table 10). This order of coefficients of variation ($PM_{2.5}$ > TVOC > CO_2) was consistent for all three seasons (not shown in results). Lower autocorrelation and higher variability of $PM_{2.5}$ samples should result in less representative samples compared to TVOC and CO_2 samples, using Luoma & Batterman's inference. However, $PM_{2.5}$ samples typically required shorter sampling periods to be considered representative in two of the three seasons evaluated in our study, compared to TVOC and CO_2 samples (Figure 26). Therefore, the reason for the variability in representativeness between IEQ indicators needs further analysis before conclusions are made.

In the heating season, higher proportions of TVOC and CO_2 samples were representative compared to samples of $PM_{2.5}$ for each analyzed sampling length (1-day, 3-day, 7-day, and 14-day) (Figure 26). The proportion of representative TVOC samples varied less between season than the proportions of representative $PM_{2.5}$ and CO_2 samples. One possible interpretation of this result is that behaviors contributing to indoor VOC generation (e.g., cleaning, clothes washing, and use of personal care products) were less variable between seasons than behaviors generating $PM_{2.5}$ (e.g., cooking). A participant may clean at consistent intervals regardless of the season, while a participant may alter cooking behaviors depending on the season (e.g., using the oven less in the cooling season to avoid heating the home). A lower proportion of $PM_{2.5}$ samples were representative in the heating season compared to other seasons, while the opposite was observed for CO_2 samples. One possible interpretation of this result is that occupant schedules were more consistent in the heating season than other seasons. Occupants may have been indoors more frequently when outdoor temperatures were lower, leaving the home only at times that are typical for the specific occupant. As CO_2 concentration is often used as a proxy for occupant presence in buildings (Wei et al., 2020) consistency in occupant behavior would likely have caused low between-day variability in diel structures of CO_2 concentrations. A final interpretation of between-season variability in representativeness considers impacts that wildfires may have had on the study. $PM_{2.5}$ concentrations recorded during the fire season were generally higher than other months (Figure 21). Outdoor $PM_{2.5}$ concentrations during the fire season may have substantially increased the average indoor hour-of-day concentrations for the six- to ten-month monitoring period. Wildfires were most active in shoulder and cooling seasons in the current study, meaning high $PM_{2.5}$ concentrations in the cooling and shoulder seasons (Figure 15) may have had greater weight in calculating the long-term average compared to concentrations in the heating season. This unequal weighting between seasons may have caused the long-term period to yield $PM_{2.5}$ diel peaks that were more similar in timing and proportion to those yielded from cooling and shoulder season samples (compared to heating season samples).

Examples comparing sample diel structures with their corresponding values of time-structured representativeness (Figure 25) suggested time-structured representativeness may be used to assess the representativeness of peak timing and proportion. For example, a sample that yields average diel peaks at times that contrast to the diel peaks yielded from the long-term monitoring period (e.g., the 3-day sample in Figure 25), will likely have a lower representativeness value than a sample with diel peaks similar in timing to those yielded in the long-term period (e.g., the 14-day sample in Figure 25). However, the inclusion of magnitude-based representativeness would be helpful to evaluate if the peaks are much higher or lower than the long-term. In the previous example, the 3-day sample yielded peaks that were more similar in magnitude to those of the long-term period than the 7-day sample peaks, but the 7-day sample was still considered more representative. Time-structured analysis alone may still be helpful to determine how long one must monitor to understand the times of day IEQ indicators yield extreme values (regardless of the exact magnitude). Energy efficiency practitioners could use information regarding hour-of-day peak timing to gain insight on possible indoor behaviors and activities that may influence daily concentration peaks, especially if other complementary high-resolution behavioral or energy use data is available.

The values of temporal representativeness in this study were not expected to be similar to those in Maciejewska et al.'s study (2015). Maciejewska et al. established 0.9 as a threshold value (i.e., the value of representativeness above which a sample was considered representative), while the we interpreted the results of the current study based on a threshold value of 0.8. We also expected a difference between study results because the primary long-term monitoring period of interest in Maciejewska et al.'s study was one month in length; the long-term period was six- to ten- months in the current study. When Maciejewska et al. calculated representativeness of IEQ samples with respect to long-term periods of varying length, they concluded the sample length needed to be approximately 20 to 50% the length of the long-term period of interest to be representative. However, the values of representativeness calculated in their study were not compared to the actual diel structure of the IEQ data. In our study, we visually inspected average diel structures of IEQ samples that varied in length, and we paired each structure with its corresponding representativeness value (Figure 25). Even with a short IEQ sample length of 3 days (representativeness calculated at 0.84, 0.79, and 0.72 for $PM_{2.5}$, TVOC and

CO_2 respectively), average hour-of-day peak values yielded from a sample were similar in timing to those yielded from the long-term period. Therefore, with the current dataset, we determined 0.8 may be a more valid representativeness threshold than 0.9.

High percentages of 3-day $PM_{2.5}$ and TVOC samples were considered representative according to the 0.8 representativeness threshold (80 to 91% of samples, depending on the season for both IEQ indicators) (Figure 26). This result suggests energy efficiency practitioners can be confident in the time of day at which $PM_{2.5}$ or TVOC peaks occur on a "typical" day based on 3-day samples; CO_2 samples may require longer lengths (4- to 7- days). Even if resources are only available to sample for one day, our analysis suggests the time structure of a $PM_{2.5}$ sample has a high likelihood of being representative. Over 75% of 1-day $PM_{2.5}$ cooling and shoulder season samples were considered representative. One-day samples of TVOC were less likely to be representative than $PM_{2.5}$ samples, but over 50% still exceeded the threshold in all seasons. Representativeness of $PM_{2.5}$ and CO_2 samples in the current study varied by season. We could not determine whether to attribute this phenomenon to IEQ indicator characteristics, occupant behavior, seasonal wildfire conditions, or a combination of these factors. Regardless, the sampling season may impact the confidence a practitioner has in the representativeness of their IEQ sample.

6.4. Spatial Representativeness and Specificity

Past studies have evaluated air quality monitoring networks via mutual information, the derived form of relative entropy that was the basis for our spatial specificity index (Osses et al., 2013; Perez-Abreu & Rodriguez, 1996; Silva & Quiroz, 2003). These studies aimed to determine the ideal combination of outdoor monitoring stations by identifying the network configuration that would lose the most information if one of the included sensors was removed. Essentially, the authors wanted to determine which sensor was the most indispensable to the network. In the current study, bedroom sensors were ranked highest in specificity and lowest in representativeness in most homes considering three indoor rooms (i.e., bedroom, living room, and kitchen) (Figure 28 and Figure 29). This result was likely observed because most of the data recorded within each home were gathered from two rooms that behaved similarly: the kitchen and the living room. As noted in a study by Wan et al. (2011), living room values of $PM_{2.5}$ values can be closely tied to kitchen values, increasing significantly during a cooking event and remaining elevated approximately 60 minutes following the event. Average hour-of-day $PM_{2.5}$ concentrations in kitchens and living rooms were above the room-pooled mean during typical cooking hours in many of the homes in our study (Figure 22). This phenomenon was likely due to short distances and lack of walls between living room and kitchen sensors in most homes, although these building characteristics were not recorded. In contrast, average hour-of-day $PM_{2.5}$ concentrations in bedrooms were below the room-pooled mean during typical cooking hours in many of the homes. Past studies have observed between-room variability in IEQ indicators. Suppressed levels of $PM_{2.5}$ in bedrooms relative to kitchen values were recorded in multiple studies (Pietrogrande et al., 2021; Xiang et al., 2021; Yassin et al., 2012). Pietrogrande et al. (2021) recorded overnight TVOC concentrations in bedrooms that were higher than kitchen concentrations. They also reported TVOC concentrations in kitchens that were higher than bedroom concentrations during cooking periods, although this was just on one example day. Results reported by Pietrogrande et al. (2021), Xiang et al. (2021) Yassin et al. (2012) and the current study therefore suggest energy efficiency practitioners could lose the most amount of information related to IEQ indicators by not installing a sensor in the bedroom. This lost information may not be relevant if practitioners are not concerned with bedroom conditions. In a sampling campaign, the benefits of adding highly specific information would have weighed against sampling resources that are often limited.

7. Conclusion

In this study, the spatial and temporal patterning of IEQ metrics were observed using commercial-grade real-time IEQ sensors within 15 homes in the City of Fort Collins. Cooling, shoulder, and heating seasons were defined by daily participant heating and cooling behaviors, which were measured using real-time energy use data from heating and cooling devices within the home. Homes were grouped by heating/cooling behaviors observed over the entire study period, with significant between-home variability observed in the date of cooling-to-heating season transitions.

Overall, median $PM_{2.5}$, TVOC, and CO_2 concentrations were lower in the heating season than in the cooling and shoulder seasons. Ranges of indoor $PM_{2.5}$, TVOC, and CO_2 concentrations were comparable between seasons. Indoor $PM_{2.5}$, TVOC, and CO_2 concentrations all exhibited higher hour-of-day average values in cooling days compared to heating days. This may have been due to participants maintaining closed windows and doors to conserve cool air-conditioned air or to prevent wildfire some from entering the home during fire season; however, ventilation and behavioral data was not gathered to discern such conclusions. Higher cooling day values may also have been due to increased infiltration of wildfire-generated $PM_{2.5}$, as levels of $PM_{2.5}$ increased both indoors and outdoors during the 2020 wildfire season.

Diel average trends of $PM_{2.5}$ suggested cooking activities in the kitchen were significant drivers of $PM_{2.5}$ in most homes. Average $PM_{2.5}$ concentrations increased at similar hours of the day between living rooms, kitchens, and bedrooms. Bedroom and living room evening peaks (around 6 pm) yielded lower $PM_{2.5}$ concentrations, on average, compared to kitchen evening peaks. Diel average TVOC concentrations in kitchens and living rooms displayed evening peaks that are likely either attributed to garage sources or increased indoor participant activity (e.g., cooking and cleaning). Bedroom average

TVOC and CO_2 concentrations tended to increase overnight, likely due to occupant presence. Correlations between $PM_{2.5}$ hourly concentrations recorded in the garage and those recorded in indoor rooms were observed to vary with a predictable pattern throughout the day. In many homes, correlations between garage $PM_{2.5}$ concentrations and living room $PM_{2.5}$ concentrations were higher from midnight to 6 am compared to correlations during daytime hours. If future studies investigated drivers and determinants of this relationship, we may expect to discover more on the mechanisms of infiltration of $PM_{2.5}$ and other pollutants from garages and outdoor areas into living spaces. Such discoveries could underscore the importance of home energy efficiency upgrades (e.g., sealing leaks in the home envelope) to improve occupant health.

The extent to which in-home IEQ samples (sampling periods ranging from one to fourteen days) represented IEQ conditions over a long-term (six- to ten- month) period was evaluated using a measure of temporal representativeness. A sample was considered representative if it exceeded a temporal representativeness value of 0.8. Temporal representativeness increased with sample length. Over 75% of 1-day $PM_{2.5}$ cooling and shoulder season samples were considered representative. Samples of TVOC that were 1-day in length were less likely to be representative, but over 50% of 1-day TVOC samples still exceeded the 0.8 threshold in all seasons. Depending on the season, 80 to 91% of 3-day samples were considered representative for $PM_{2.5}$ and TVOC samples. In cooling and shoulder seasons, higher proportions of $PM_{2.5}$ samples were representative compared to those of TVOC, and higher proportions of TVOC samples were representative compared to those of CO_2. Representativeness of $PM_{2.5}$ and CO_2 samples varied by season. These results suggest energy efficiency practitioners can be confident in the time of day at which $PM_{2.5}$ or TVOC peaks occur on a “typical” day, based on 3-day samples; CO_2 samples may require longer lengths. Even if resources are only available to sample for one day, our analysis suggests the time structure of a $PM_{2.5}$ sample (i.e., the hour(s) when concentration peaks during the day) has a high likelihood of being representative of a “typical” day, although this may vary depending on the season of the sample.

Spatial representativeness and the complimentary measure of spatial specificity were evaluated for in-home IEQ samples. In most homes, $PM_{2.5}$, TVOC, and CO_2 data gathered from bedrooms were the most specific (i.e., yielded the highest specificity values) out of all three considered rooms (bedroom, living room, and kitchen). Essentially, the dataset collected in the bedroom for these three IEQ metrics was the most unique out of the datasets collected from the three considered rooms. This suggests that if energy efficiency practitioners are aiming to observe the full range of IEQ variability between living spaces, and they are only able to install IEQ sensors in two rooms within a home, the bedroom should be one of the rooms sampled. However, the data gathered from the bedroom may only be applicable if conditions within the bedroom are of interest; bedroom IEQ conditions may not be applicable to other living areas within the home. The living room and kitchen were likely to report similar trends and magnitudes in $PM_{2.5}$, TVOC, and CO_2 data. Relationships between room, spatial representativeness/specificity, and other variables, such as distance between rooms and HVAC structure, should be explored to discover why between-room variability is higher in certain homes.

This study proposed quality criteria that may be used to evaluate fit-for-purpose among IEQ indicators in frameworks developed to assess residential energy efficiency upgrade impacts. In particular, we focused on the measures of temporal representativeness, spatial representativeness, and spatial specificity. These three measures may suggest the temporal and spatial resolution at which IEQ indicators should be measured to adequately characterize the typical long-term conditions within a home. Findings have implications for those aiming to develop best practices when taking short samples of IEQ indicators in homes, whether they be energy efficiency practitioners determining the impacts of residential upgrades or researchers considering IEQ impacts on occupant health.

8. Limitations

While presenting a unique possibility for data collection, the COVID-19 pandemic and wildfires that occurred during the study period may have caused the behaviors and measured IEQ indicators to be less representative of “typical” conditions within Fort Collins homes. Stay-at-home orders established at the start of the study period in the summer and fall months of 2020 caused more occupants to work from home. These orders likely increased general in-home activity above levels that were typical before the pandemic. Wildfires caused an increase in $PM_{2.5}$ levels and may have caused participants to keep windows shut more than usual between August 2020 and October 2020. Both these anomalies may suggest results are less generalizable and less comparable to those reported in other IEQ studies.

Seasons defined in this study did not occur during the same time periods between homes. Therefore, some variables that were present in the shoulder seasons of some homes (e.g., wildfires) may not have been present in the shoulder seasons of others. Between-home comparison disaggregated by season was not common, so this was assumed to be an insignificant issue in analysis. Seasons also varied in length. This may have impacted the representativeness of samples gathered in one season more than the representativeness of samples gathered in another. Although the majority of data gathered in most homes was gathered in the heating season, $PM_{2.5}$ samples from this season were, on average, less representative compared to samples from other seasons. It was expected that heating days would be most representative of the long-term (six- to ten- month) monitoring period, if most of the long-term period consisted of heating days. A longer season resulting in less representative samples suggests the number of samples within each season was sufficiently high for the difference in sample number to have little effect. However, no extensive analysis was performed to reach this conclusion.

TVOC, one of the IEQ indicators considered most in this study, is an aggregate response to many compounds that come from various sources and have varying impacts on human health. While this IEQ indicator may give an idea of general trends in occupant or resident activities, conclusions could not be made on what caused values to increase/decrease or what the trends implied for residents.

The home energy monitor has not, to our knowledge, been used in many past studies. We therefore considered the monitor experimental. Generally, as monthly natural gas usage increased, so did the amount of time on for the primary heating device identified by the energy monitor in the homes for which natural gas data was available. However, natural gas data for some homes showed substantial natural gas consumption in October and November when the primary heating device was not detected. This discrepancy suggests that primary heating device usage identified by the energy monitor may not have been a reliable proxy in all months for all homes. Lack of primary heating device identification may have caused days to be classified as cooling or shoulder days even though occupants were using high levels of space heating. This in turn, could have caused some days to be grouped incorrectly in our analysis.

It was not desirable to make assumptions about which devices were identified incorrectly by the home energy monitor. Some participants may use air conditioning in months that do not seem "typical." Therefore, appliances defined by the home energy monitor were, for the most part assumed to be correctly classified. Examples of home energy monitor identification that were uncertain were the primary heating devices identified for Homes 1 and 11. December 3rd, 2020 was the first day on which the primary heating device was detected for both of these homes, despite both homes' enrollment date occurring months prior. This simultaneous date of device identification may have been due to the home energy monitor experiencing an update that allowed it to detect additional devices. Similarly, the primary heating device identified for Home 10 recorded a period of apparent inactivity from October 19th, 2020 to December 3rd, 2020. This could have been due to a period of home vacancy (e.g., the residents shut primary heat off while traveling), but the date the device was identified once more (December 3rd – i.e., the same as Homes 1 and 11) is possible cause for concern. If the categorization of these days were artifacts of home energy monitor misidentification, this could have caused data from days that should have been defined as heating days to be pooled in shoulder day data instead.

We depended on participants to inform us when they received upgrades or major changes occurred to their homes. We administered questionnaires partway through the study asking about upgrades, and few participants had acquired any significant upgrades to their homes. However, some participants may have switched out appliances without our knowledge. This may have led to changes in the home energy monitor data partway through the study that were not based on behavioral changes, but technological modifications.

The statistical measure of relative entropy was used to define a numerical value of temporal representativeness in this study. Home IEQ samples in this study (ranging from one to 28 days in length) were assigned numerical representativeness values (ranging from 0 to 1) that quantified how well the samples represented home IEQ conditions over the long-term monitoring period (six to ten months). A representativeness threshold value of 0.8 was defined, meaning any sample with a representativeness above 0.8 was considered representative. Recommendations of the minimum lengths required for representative sampling periods (e.g., one day vs one week) were made upon the basis of this threshold. However, conclusions on how well sampling periods of different lengths characterize long-term periods cannot be made with certainty without further exploration of relative entropy as a measure of representativeness.

Our measures of spatial representativeness and spatial specificity were adapted from a study of outdoor air quality monitoring stations performed by Osses et al. (2013). In their study, Osses et al. calculated representativity (from which spatial representativeness in our study was adapted) and specificity of each monitoring station in a monitoring network. The monitoring network considered in their study consisted of seven monitoring stations spread across the city of Santiago, Chile. In our study, we calculated spatial representativeness and spatial specificity for indoor environmental quality (IEQ) indicator data recorded in three indoor living areas (i.e., the living room, the kitchen, and the bedroom) in each home. The methods used to calculate representativity of a station in Osses et al.'s study and spatial representativeness of a room in our study relied upon the creation of a modeled dataset, which would predict air quality or IEQ indicator values if data from the monitor (or room) of interest were not available. A high representativity (spatial representativeness in our study) value is calculated for a sensor if adding the sensor to the network greatly decreases the uncertainty in the model. Osses et al. used a priori data (air quality data for the City of Santiago collected prior to their study period) to model the air quality of Santiago without including data from the monitor of interest. No a priori data was available in our study from within the recruited homes. Our study instead simply combined the data from the two "other" rooms (i.e., not the room of interest) to create a data distribution that would act as the "modeled" data distribution in the calculation of spatial representativeness. Our method limits the applicability of the measure of spatial representativeness used in this study, as our method guaranteed that the room with data that differed the most from the other two rooms in a home would be assigned a lower value of spatial representativeness. This essentially caused spatial representativeness to decrease whenever spatial specificity increased, making one of the two measures redundant. Future studies aiming to calculate spatial representativeness of IEQ sensors compared between multiple rooms could use characteristics collected from each home (e.g., distance between rooms/sensors, occupant activity patterns, air exchange rate, outdoor data) to create modeled datasets. Spatial representativeness could therefore be calculated using a dataset that is not directly dependent on the other two rooms. This would allow spatial representativeness to quantify how well data from the considered room characterizes modeled results of overall home IEQ, instead of quantifying how well data from the considered room characterizes IEQ within the other two rooms (which is what this study's measure of spatial representativeness essentially measured).

Future Research

These data and those of similar spatial and temporal resolution may be leveraged in a multitude of different directions. Sleep studies could use the CO_2 data in bedrooms to observe building structure and resident behavioral impacts on CO_2 levels and corresponding health effects. Source apportionment could be conducted in enrolled homes to give more certain conclusions on the between-room and between-garage/indoor correspondences observed in this study. Periods of vacancy could be identified post-hoc, and data could be aggregated by vacant and occupied periods to explore how residents impact the indoor environment through paired analysis.

Future work with the data collected in this study (or studies of similarly high temporal resolution) could further assess the correspondence between energy use metrics and IEQ indicators on a temporal scale. An example would be to observe how diel patterns in stove energy use correlated with $PM_{2.5}$ and TVOC levels. Also, as the Epic Homes program is ongoing, the addition of more homes to the dataset would provide opportunities to evaluate relationships between IEQ indicators and energy use metrics. Regression models that include IEQ indicators, energy use metrics, home characteristics, and resident behaviors could be developed. This would also allow the responsiveness of IEQ metrics (one of the quality criteria proposed in our study) to be evaluated through changes in IEQ and energy use variables pre- and post-upgrade. These types of analyses could give practitioners and researchers a better understanding of the causes of inadequate home IEQ.

Representativeness of other IEQ indicators and energy use metrics could be further explored to determine how the required sampling spatial and temporal resolutions vary by indicator. Also, as mentioned in the discussion on limitations, defining thresholds of representativeness may improve standardization of best practices for IEQ monitoring in both residential and non-residential settings. Best practices could mature to include recommended sampling durations and locations within homes (which may vary depending on the aim of those gathering data). Once data from more homes are collected, or as other studies gather data with similar resolution in larger samples of homes, the impact of home structure (e.g., air exchange rate) and resident characteristics and activities (e.g., housing tenure; socioeconomic position) on representativeness could be explored.

Lastly, as previously noted, a significant portion of the data collection for this study occurred when many participants were working from home due to COVID-19 lockdowns. While this may be a limitation to the study’s generalizability, researchers may benefit from comparing this dataset with other high-resolution datasets collected during “typical” periods when most residents were less likely to work from home.

References

Abdullahi, K. L., Delgado-Saborit, J. M., & Harrison, R. M. (2013). Emissions and indoor concentrations of particulate matter and its specific chemical components from cooking: A review. *Atmospheric Environment*, *71*, 260–294. https://doi.org/10.1016/j.atmosenv.2013.01.061

Afshar-Mohajer, N., Foos, R., Volckens, J., & Ramachandran, G. (2020). Variability of aerosol mass and number concentrations during taconite mining operations. *Journal of Occupational and Environmental Hygiene*, *17*(1), 1–14. https://doi.org/10.1080/15459624.2019.1688823

Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, *88*(17). https://doi.org/https://doi.org/10.1103/PhysRevLett.88.174102

Bari, A., Kindzierski, W. B., Wheeler, A. J., Eve, M.-, & Wallace, L. A. (2015). Source apportionment of indoor and outdoor volatile organic compounds at homes in Edmonton, Canada. *Building and Environment*, *90*. https://doi.org/10.1016/j.buildenv.2015.03.023

Basu, C., Pau, V. K., & Syal, M. G. M. (2019). Performance indicators for energy efficiency retrofitting in multifamily residential buildings. *Journal of Green Building*, *14*(2), 109–136. https://doi.org/10.3992/1943-4618.14.2.109

Batterman, S., Jia, C., & Hatzivasilis, G. (2007). Migration of volatile organic compounds from attached garages to residences : A major exposure source. *Environmental Research*, *104*, 224–240. https://doi.org/10.1016/j.envres.2007.01.008

Becchio, C., Bottero, M. C., Corgnati, S. P., & Dell'Anna, F. (2018). Decision making for sustainable urban energy planning: an integrated evaluation framework of alternative solutions for a NZED (Net Zero-

Energy District) in Turin. *Land Use Policy*, *78*(February), 803–817. https://doi.org/10.1016/j.landusepol.2018.06.048

Bekö, G., Gustavsen, S., Frederiksen, M., Bergsøe, N. C., Kolarik, B., Gunnarsen, L., Toftum, J., & Clausen, G. (2016). Diurnal and seasonal variation in air exchange rates and interzonal airflows measured by active and passive tracer gas in homes. *Building and Environment*, *104*, 178–187. https://doi.org/10.1016/j.buildenv.2016.05.016

Bousquet, N. (2008). *Diagnostics of prior-data agreement in applied Bayesian analysis Diagnostics of prior-data agreement in*. *4763*. https://doi.org/10.1080/02664760802192981

Brockwell, P. J., & Davis, R. A. (2016). *Springer Texts in Statistics Introduction to Time Series and Forecasting* (3rd ed.). Springer International Publishing. http://www.springer.com/series/417

Building Research Establishment Ltd. (2022). *BREEAM*. https://www.breeam.com/

Caselton, W. F., & Zidek, J. V. (1984). Optimal monitoring network designs. *Statistics & Probability Letters*, *2*, 223–227. https://doi.org/https://doi.org/10.1016/0167-7152(84)90020-8

Certivéa. (2022). *OsmoZ*. https://osmoz.certivea.fr/

Clarisse, B., Laurent, A. M., Seta, N., le Moullec, Y., el Hasnaoui, A., & Momas, I. (2003). Indoor aldehydes: Measurement of contamination levels and identification of their determinants in Paris dwellings. *Environmental Research*, *92*(3), 245–253. https://doi.org/10.1016/S0013-9351(03)00039-2

Clements, N., Keady, P., Emerson, J. B., Fierer, N., & Miller, S. L. (2018). Seasonal Variability of Airborne Particulate Matter and Bacterial Concentrations in Colorado Homes. *Atmosphere*, *9*(133), 1–18. https://doi.org/10.3390/atmos9040133

Clements, N., Zhang, R., Jamrozik, A., Campanella, C., & Bauer, B. (2019). The spatial and temporal variability of the indoor environmental quality during three simulated office studies at a living lab. *Buildings*, *9*(62). https://doi.org/10.3390/buildings9030062

Colorado Department of Public Health & Environment. (2019). *2019 Air Quality Data Report*.

Das, P., Chalabi, Z., Jones, B., Milner, J., Shrubsole, C., Davies, M., Hamilton, I., Ridley, I., & Wilkinson, P. (2013). Multi-objective methods for determining optimal ventilation rates in dwellings. *Building and Environment*, *66*, 72–81. https://doi.org/10.1016/j.buildenv.2013.03.021

Dellinger, A. B., & Leech, N. L. (2007). Toward a Unified Validation Framework in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(4), 309–332. https://doi.org/10.1177/1558689807306147

Deng, S., Zou, B., & Lau, J. (2021). The Adverse Associations of Classrooms' Indoor Air Quality and Thermal Comfort Conditions on Students' Illness Related Absenteeism between Heating and Non-Heating Seasons-A Pilot Study. *Public Health*, *18*. https://doi.org/10.3390/ijerph

Diaz Lozano Patino, E., & Siegel, J. A. (2018). Indoor environmental quality in social housing: A literature review. *Building and Environment*, *131*(January), 231–241. https://doi.org/10.1016/j.buildenv.2018.01.013

Dodson, R. E., Levy, J. I., Spengler, J. D., Shine, J. P., & Bennett, D. H. (2008). Influence of basements, garages, and common hallways on indoor residential volatile organic compound concentrations. *Atmospheric Environment*, *42*, 1569–1581. https://doi.org/10.1016/j.atmosenv.2007.10.088

Du, J., Pan, W., & Yu, C. (2020). In-situ monitoring of occupant behavior in residential buildings – a timely review. *Energy & Buildings*, *212*. https://doi.org/10.1016/j.enbuild.2020.109811

E4 the Future. (2016). *Occupant Health Benefits of Residential Energy Efficiency*. https://e4thefuture.org/wp-content/uploads/2016/11/Occupant-Health-Benefits-Residential-EE.pdf

Escobedo, L. E., Champion, W. M., Li, N., & Montoya, L. D. (2014). Indoor air quality in Latino homes in Boulder, Colorado. *Atmospheric Environment*, *92*, 69–75. https://doi.org/10.1016/j.atmosenv.2014.03.043

Ezratty, V., Ormandy, D., Laurent, M. H., Boutière, F., Duburcq, A., Courouve, L., & Cabanes, P. A. (2018). Assessment of the health-related costs and benefits of upgrading energy efficiency in French housing. *Environnement, Risques et Sante*, *17*(4), 401–410. https://doi.org/10.1684/ers.2018.1200

Fazli, T., & Stephens, B. (2018). Development of a nationally representative set of combined building energy and indoor air quality models for U.S. residences. *Building and Environment*, *136*, 198–212. https://doi.org/10.1016/j.buildenv.2018.03.047

Fisher Sheehan & Colton. (2020). Home Energy Affordability Gap. *Public Finance and General Economics*.

Fregonara, E., & Pattono, S. (2018). A sustainability indicator for building projects in presence of risk/uncertainty over time: A research experience. *Aestimum*, *2018*(73), 173–205. https://doi.org/10.13128/Aestimum-24928

Fugler, D., Grande, C., & Graham, L. (2002). Attached garages are likely path for pollutants. *ASHRAE IAQ Applications*, *3*.

Gallagher, M., Wysocki, C. J., Leyden, J. J., Spielman, A. I., Sun, X., & Preti, G. (2008). Analyses of volatile organic compounds from human skin. *British Journal of Dermatology*, *159*(4), 780–791. https://doi.org/10.1111/j.1365-2133.2008.08748.x

Gerbing, D. (2016). *Autocorrelation*. School of Business Administration, Portland State Uinversity.

Graham, L., O'Leary, K., & Noseworthy, L. (1999). *Indoor air sampling for infiltration of vehicle emissions to the house from an attached garage, Report #99-26768.*

Hayashi, P., Abib, G., & Hoppen, N. (2019). Validity in qualitative research: A processual approach. *Qualitative Report*, *24*(1), 98–112. https://doi.org/10.46743/2160-3715/2019.3443

Hayes, S., Kubes, C., & Gerbode, C. (2020). Making Health Count: Monetizing the Health Benefits of In-Home Services Delivered by Energy Efficiency Programs. *American Council for an Energy Efficient Environment*.

Héroux, M., Clark, N., Ryswyk, K. van, Mallick, R., Anastassopoulos, A., Guay, M., Macneill, M., & Wheeler, A. J. (2010). Predictors of Indoor Air Concentrations in Smoking and Non-Smoking Residences. *International Journal of Environmental Research and Public Health*, *7*, 3080–3099. https://doi.org/10.3390/ijerph7083080

Hollnagel, E. (2014). Resilience engineering and the built environment. *Building Research & Information*, *42*(2), 221–228. https://doi.org/https://doi.org/10.1080/09613218.2014.862607

Holton, C., Luo, H., Dahlen, P., Gorder, K., Dettenmaier, E., & Johnson, P. C. (2013). Temporal variability of indoor air concentrations under natural conditions in a house overlying a dilute chlorinated solvent groundwater plume. *Environmental Science and Technology*, *47*(23), 13347–13354. https://doi.org/10.1021/es4024767

Institue Valencià de l'Edificació. (2022). *Certificatión de Oficinas – BES Bienestar en Espacios Sostenibles*. https://www.five.es/oficinas/

International Energy Agency. (2014). Capturing the Multiple Benefits of Energy Efficiency. *Capturing the Multiple Benefits of Energy Efficiency*. https://doi.org/10.1787/9789264220720-en

International Energy Agency. (2017). *Indoor Air Quality Design and Control in Low-energy Residential Buildings- Annex 68 | Subtask 1: Defining the metrics*. *September*.

International WELL Building Institute. (2022). *WELL Health-Safety Rating*. International WELL Building Institute

Jaffe, D., Hafner, W., Chand, D., Westerling, A., & Spracklen, D. (2008). Interannual variations in PM2.5 due to wildfires in the Western United States. *Environmental Science and Technology*, *42*(8), 2812–2818. https://doi.org/10.1021/es702755v

Jia, C., Cao, K., Valaulikar, R., & Fu, X. (2019). Variability of Total Volatile Organic Compounds (TVOC) in the Indoor Air of Retail Stores. *International Journal of Environmental Research and Public Health*. https://doi.org/https://doi.org/10.3390/ijerph16234622

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. v, Hern, S. C., & Engelmann, W. H. (2001). The National Human Activity Pattern Survey. *Journal of Exposure Analysis and Environmental Epidemiology*, *11*(3), 231–252. https://doi.org/https://doi.org/10.1038/sj.jea.7500165

Kowalski, A., Martin, M. T., Plastino, A., & Judge, G. (2014). Kullback-Leibler Approach to Chaotic Time Series. *SOP Transactions on Theoretical Physics*, *1*(October). https://doi.org/10.15764/TPHY.2014.03003

Kullback, S. (1959). *Information Theory and Statistics* (1st ed.). Wiley.

Lai, S. C., Ho, K. F., Zhang, Y. Y., Lee, S. C., Huang, Y., & Zou, S. C. (2010). Characteristics of residential indoor carbonaceous aerosols: A case study in Guangzhou, Pearl river delta region. *Aerosol and Air Quality Research*, *10*(5), 472–478. https://doi.org/10.4209/aaqr.2010.05.0037

Lee, E. G., Feigley, C. E., Hussey, R., & Slaven, J. E. (2008). Estimation of required monitoring time for obtaining validation data in enclosed spaces. *Journal of Environmental Monitoring*, 1350–1356. https://doi.org/10.1039/b806421k

Lee, K., Bartell, S. M., & Paek, D. (2004). Interpersonal and daily variability of personal exposures to nitrogen dioxide and sulfur dioxide. *Journal of Exposure Analysis and Environmental Epidemiology*, *14*, 137–143. https://doi.org/10.1038/sj.jea.7500304

Luoma, M., & Batterman, S. A. (2000). Autocorrelation and Variability of Indoor Air Quality Measurements. *American Industrial Hygiene Association Journal*, *61*. https://doi.org/10.1080/15298660008984575

Maciejewska, M., & Szczurek, A. (2015). Representativeness of shorter measurement sessions in long-term indoor air monitoring. *Environmental Science Processes & Impacts*, 381–388. https://doi.org/10.1039/c4em00409d

Mallach, G., Aubin, M. M. D., Shin, L. W. T., Kulka, K. V. R. R., Fugler, H. Y. D., & Wheeler, E. L. A. J. (2017). Exhaust ventilation in attached garages improves residential indoor air quality. *Indoor Air*, *27*(July 2016), 487–499. https://doi.org/10.1111/ina.12321

Massey, D., Kulshrestha, A., Masih, J., & Taneja, A. (2012). Seasonal trends of PM10, PM5.0, PM2.5 & PM1.0 in indoor and outdoor environments of residential homes located in North-Central India. *Building and Environment*, *47*(1), 223–231. https://doi.org/10.1016/j.buildenv.2011.07.018

Matz, C. J., Stieb, D. M., Davis, K., Egyed, M., Rose, A., Chou, B., & Brion, O. (2014). Effects of age, season, gender and urban-rural status on time-activity: Canadian human activity pattern survey 2 (CHAPS 2). *International Journal of Environmental Research and Public Health*, *11*(2), 2108–2124. https://doi.org/10.3390/ijerph110202108

Militello-Hourigan, R. E., & Miller, S. L. (2018). The impacts of cooking and an assessment of indoor air quality in Colorado passive and tightly constructed homes. *Building and Environment*, *144*(June), 573–582. https://doi.org/10.1016/j.buildenv.2018.08.044

Miller, S. L., Scaramella, P., Campe, J., Goss, C. W., Diaz-Castillo, S., Hendrikson, E., DiGuiseppi, C., & Litt, J. (2009). An assessment of indoor air quality in recent Mexican immigrant housing in Commerce City, Colorado. *Atmospheric Environment*, *43*(35), 5661–5667. https://doi.org/10.1016/j.atmosenv.2009.07.037

Molhave, L., Clausen, G., Berglund, B., Ceaurriz, J. de, Kettrup, A., Lindvall, T., Maron, M., Pickering, A., Rothweiler, H., Seifert, B., & Ceaurriz, de. (1997). Total Volatile Organic Compounds (TVOC) in Indoor Air Quality Investigations. *Indoor Air*, *7*, 225–240. https://doi.org/10.1111/j.1600-0668.1997.00002.x

Oak Ridge National Laboratory. (1990). *1990 National Evaluation of the Weatherization Assistance Program: Summary of Final Reports*. https://weatherization.ornl.gov/wp-content/uploads/pdf/WAP1989/1990_final_report_summary.pdf

Ortiz, J., Casquero-Modrego, N., & Salom, J. (2019). Health and related economic effects of residential energy retrofitting in Spain. *Energy Policy*, *130*(March), 375–388. https://doi.org/10.1016/j.enpol.2019.04.013

Osses, A., Gallardo, L., & Faundez, T. (2013). Analysis and evolution of air quality monitoring networks using combined statistical information indexes. *Tellus, Series B: Chemical and Physical Meteorology*, *65*. https://doi.org/10.3402/tellusb.v65i0.19822

Perez-Abreu, V., & Rodriguez, J. E. (1996). Index of effectiveness of a multivariate environmental monitoring network. *Environmetrics*, *7*(April), 489–501.

https://doi.org/https://doi.org/10.1002/(SICI)1099-095X(199609)7:5%3C489::AID-ENV224%3E3.0.CO;2-0

Pietrogrande, M. C., Casari, L., Demaria, G., & Russo, M. (2021). Indoor Air Quality in Domestic Environments during Periods Close to Italian COVID-19 Lockdown. *International Journal of Environmental Research and Public Health*, *18*(4060). https://doi.org/https://doi.org/10.3390/ijerph18084060

Pombo, O., Rivela, B., & Neila, J. (2016). The challenge of sustainable building renovation: Assessment of current criteria and future outlook. *Journal of Cleaner Production*, *123*, 88–100. https://doi.org/10.1016/j.jclepro.2015.06.137

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rösch, C., Kohajda, T., Röder, S., Bergen, M. von, Schlink, U., & Rösch, C. (2014). Relationship between sources and patterns of VOCs in indoor air. *Atmospheric Pollution Research*, *5*, 129–137. https://doi.org/10.5094/APR.2014.016

SAGE. (2017). *Learn About Time Series ACF and PACF in SPSS With Data From the USDA Feed Grains Database*. SAGE Publications Ltd. https://methods.sagepub.com/base/download/DatasetStudentGuide/time-series-acf-pacf-in-us-feedgrains-1876-2015#:~:text=A%20PACF%20is%20similar%20to,points%20at%20time%20t%20%E2%88%92%201.

Seifert, B. (1995). Validity criteria for exposure assessment methods. *Science of the Total Environment*, *168*(2), 101–107. https://doi.org/10.1016/0048-9697(95)04613-6

Sensirion. (2019). *Total Volatile Organic Compounds (TVOC) and Indoor Air Quality (IAQ): SGP30 TVOC and CO2eq Sensor* (pp. 990–1005). https://doi.org/10.1007/s00103-007-0290-y

Silva, C., & Quiroz, A. (2003). Optimization of the atmospheric pollution monitoring network at Santiago de Chile. *Atmospheric Environment*, *37*, 2337–2345. https://doi.org/10.1016/S1352-2310(03)00152-3

Stanley, K., Yoo, E. H., Paul, T., & Bell, S. (2018). How many days are enough?: capturing routine human mobility. *International Journal of Geographical Information Science*, *32*(7), 1485–1504. https://doi.org/10.1080/13658816.2018.1434888

StatSoft. (2011). Tiime Series Analysis. In *Electronic Statistics Textbook*. https://www.betterevaluation.org/en/evaluation-options/timeseriesanalysis

Szigeti, T., Dunster, C., Cattaneo, A., Spinazzè, A., Mandin, C., Le Ponner, E., de Oliveira Fernandes, E., Ventura, G., Saraga, D. E., Sakellaris, I. A., de Kluizenaar, Y., Cornelissen, E., Bartzis, J. G., & Kelly, F. J. (2017). Spatial and temporal variation of particulate matter characteristics within office buildings — The OFFICAIR study. *Science of the Total Environment*, *587–588*, 59–67. https://doi.org/10.1016/j.scitotenv.2017.01.013

Tan, C. C. L., Finney, K. N., Chen, Q., Russell, N. V., Sharifi, V. N., & Swithenbank, J. (2013). Experimental investigation of indoor air pollutants in residential buildings. *Indoor and Built Environment*, *22*(3), 471–489. https://doi.org/10.1177/1420326X12441806

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Tonn, B., Rose, E., & Hawkins, B. (2018). Evaluation of the U.S. department of energy's weatherization assistance program: Impact results. *Energy Policy*, *118*, 279–290. https://doi.org/10.1016/j.enpol.2018.03.051

Touceda, M. I., Neila, F. J., & Degrez, M. (2018). Modeling socioeconomic pathways to assess sustainability: a tailored development for housing retrofit. *International Journal of Life Cycle Assessment*, *23*(3), 710–725. https://doi.org/10.1007/s11367-016-1194-6

Tryner, J., Good, N., Wilson, A., Clark, M. L., Peel, J. L., & Volckens, J. (2019). Variation in gravimetric correction factors for nephelometer-derived estimates of personal exposure to PM2.5. *Environmental Pollution*, *250*, 251–261. https://doi.org/10.1016/j.envpol.2019.03.121

Turner, W. J. N., Logue, J. M., & Wray, C. P. (2013). A combined energy and IAQ assessment of the potential value of commissioning residential mechanical ventilation systems. *Building and Environment*, *60*, 194–201. https://doi.org/10.1016/j.buildenv.2012.10.016

Underhill, L. J., Fabian, M. P., Vermeer, K., Sandel, M., Adamkiewicz, G., Leibler, J. H., & Levy, J. I. (2018). Modeling the resiliency of energy-efficient retrofits in low-income multifamily housing. *Indoor Air*, *28*(3), 459–468. https://doi.org/10.1111/ina.12446

Underhill, L. J., Milando, C. W., Levy, J. I., Dols, W. S., Lee, S. K., & Fabian, M. P. (2020). Simulation of indoor and outdoor air quality and health impacts following installation of energy-efficient retrofits in a multifamily housing unit. *Building and Environment*, *170*(June 2019), 106507. https://doi.org/10.1016/j.buildenv.2019.106507

United States Census Bureau. (2019). *Fort Collins city, Colorado*. U.S. Census Bureau's American Community Survey Office. https://data.census.gov/cedsci/profile?g=1600000US0827425

Urso, P., Cattaneo, A., Garramone, G., Peruzzo, C., Maria, D., & Carrer, P. (2015). Identification of particulate matter determinants in residential homes. *Building and Environment*, *86*, 61–69. https://doi.org/10.1016/j.buildenv.2014.12.019

U.S. Department of Energy. (2022). *Weatherization Assistance Program*. https://www.energy.gov/eere/wap/weatherization-assistance-program

U.S. Department of Energy, & Oak Ridge National Laboratory. (2010). *Weatherization Assistance Program, National Evaluations: Summary of Results*. http://weatherization.ornl.gov.

U.S. Energy Information Administration. (2021). *Residential Energy Consumption Survey (RECS)*. https://www.eia.gov/consumption/residential/index.php

US Environmental Protection Agency. (n.d.). *Volatile Organic Compounds' Impact on Indoor Air Quality*. Retrieved October 23, 2021, from https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality

US Environmental Protection Agency. (2021). *NAAQS Table*. https://www.epa.gov/criteria-air-pollutants/naaqs-table

U.S. Green Building Council. (2018). *LEED v4.1 Operations and Maintenance*. https://www.usgbc.org/leed/v41

U.S. Green Building Council. (2021). *LEED Rating System*. https://new.usgbc.org/leed

Vardoulakis, S., Giagloglou, E., Steinle, S., Davis, A., Sleeuwenhoek, A., Galea, K. S., Dixon, K., & Crawford, J. O. (2020). Indoor exposure to selected air pollutants in the home environment: A systematic review. *International Journal of Environmental Research and Public Health*, *17*(23), 1–24. https://doi.org/10.3390/ijerph17238972

Vermont Energy Investment Corporation. (2019). *Energy-Plus-Health Playbook*. *July*. https://www.veic.org/clients-results/reports/energy-plus-health-playbook

Volckens, J., Leith, C. Q. D., & Henry, J. M. C. S. (2017). Development and evaluation of an ultrasonic personal aerosol sampler. *Indoor Air*, *27*(June 2016), 409–416. https://doi.org/10.1111/ina.12318

Wan, M. P., Wu, C. L., Sze To, G. N., Chan, T. C., & Chao, C. Y. H. (2011). Ultrafine particles, and PM2.5 generated from cooking in homes. *Atmospheric Environment*, *45*(34), 6141–6148. https://doi.org/10.1016/j.atmosenv.2011.08.036

Watson, J., Fujita, E., Chow, J. C., Zielinska, B., RIchards, L., Neff, W., & Dietrich, D. (1998). *Northern Front Range Air Quality Study Final Report. Prepared for Colorado State University, Fort Collns, CO and EPRI, Palo Alto, CA*.

Wei, W., Wargocki, P., Zirngibl, J., Bendžalová, J., & Mandin, C. (2020). Review of parameters used to assess the quality of the indoor environment in Green Building certification schemes for offices and hotels. *Energy and Buildings*, *209*, 109683. https://doi.org/10.1016/j.enbuild.2019.109683

Weiss, J. (2012). Statistics for Environmental Science: Lecture 17. In *Curriculum for the Environment and Ecology*. University of North Carolina. https://sakai.unc.edu/access/content/group/2842013b-58f5-4453-aa8d-3e01bacbfc3d/public/Ecol562_Spring2012/docs/lectures/lecture17.htm#testing

Wilson, J., Jacobs, D., Reddy, A., Tohn, E., Cohen, J., & Jacobson, E. (2016). *Home Rx: The Health Benefits of Home Performance*. https://www.energy.gov/eere/buildings/downloads/home-rx-health-benefits-home-performance-review-current-evidence

Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes*, *9*(1), 8. https://doi.org/10.1186/1477-7525-9-8

World Health Organization. (2010). *WHO Guidelines for Indoor Air Quality: Selected Pollutants*.

World Health Organization. (2018). *WHO Housing and Health Guidelines*.

World Health Organization. (2021). *WHO global air quality guidelines*.

Xiang, J., Hao, J., Austin, E., Shirai, J., & Seto, E. (2021). Residential cooking-related PM2.5: Spatial-temporal variations under various intervention scenarios. *Building and Environment*, *201*(May), 1–10. https://doi.org/10.1016/j.buildenv.2021.108002

Xu, P., Chan, E. H. W., & Lam, P. T. I. (2013). A conceptual framework for delivering sustainable building energy efficency retrofit using the energy performance contracting (EPC) in China. *Journal of Green Building*, *8*(1), 177–190. https://doi.org/10.3992/jgb.8.1.177

Yassin, M. F., AlThaqeb, B. E. Y., & Al-Mutiri, E. A. E. (2012). Assessment of indoor PM 2.5 in different residential environments. *Atmospheric Environment*, *56*, 65–68. https://doi.org/10.1016/j.atmosenv.2012.03.051

Yoo, E. hye. (2019). How Short Is Long Enough? Modeling Temporal Aspects of Human Mobility Behavior Using Mobile Phone Data. *Annals of the American Association of Geographers*, *109*(5), 1415–1432. https://doi.org/10.1080/24694452.2019.1586516

Zhao, N., Liu, Y., Vanos, J. K., & Cao, G. (2018). Day-of-week and seasonal patterns of PM2.5 concentrations over the United States: Time-series analyses using the Prophet procedure. *Atmospheric Environment*, *192*, 116–127. https://doi.org/10.1016/j.atmosenv.2018.08.050

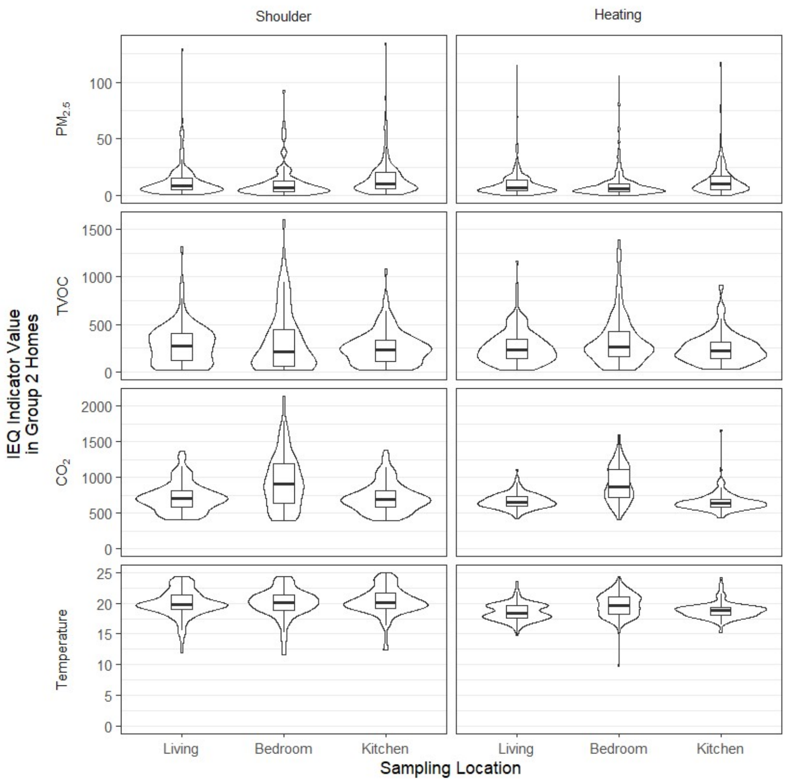


Figure 30: Distributions of time-averaged hourly IEQ parameter values from indoor rooms in Group 2 homes (homes with no definable cooling period), shown separately by season and sampling location. Values from all rooms in all Group 2 homes pooled. Units: $PM_{2.5}$ (µg/m^3), TVOC (ppb), CO_2 (ppm), Temperature (^oC).

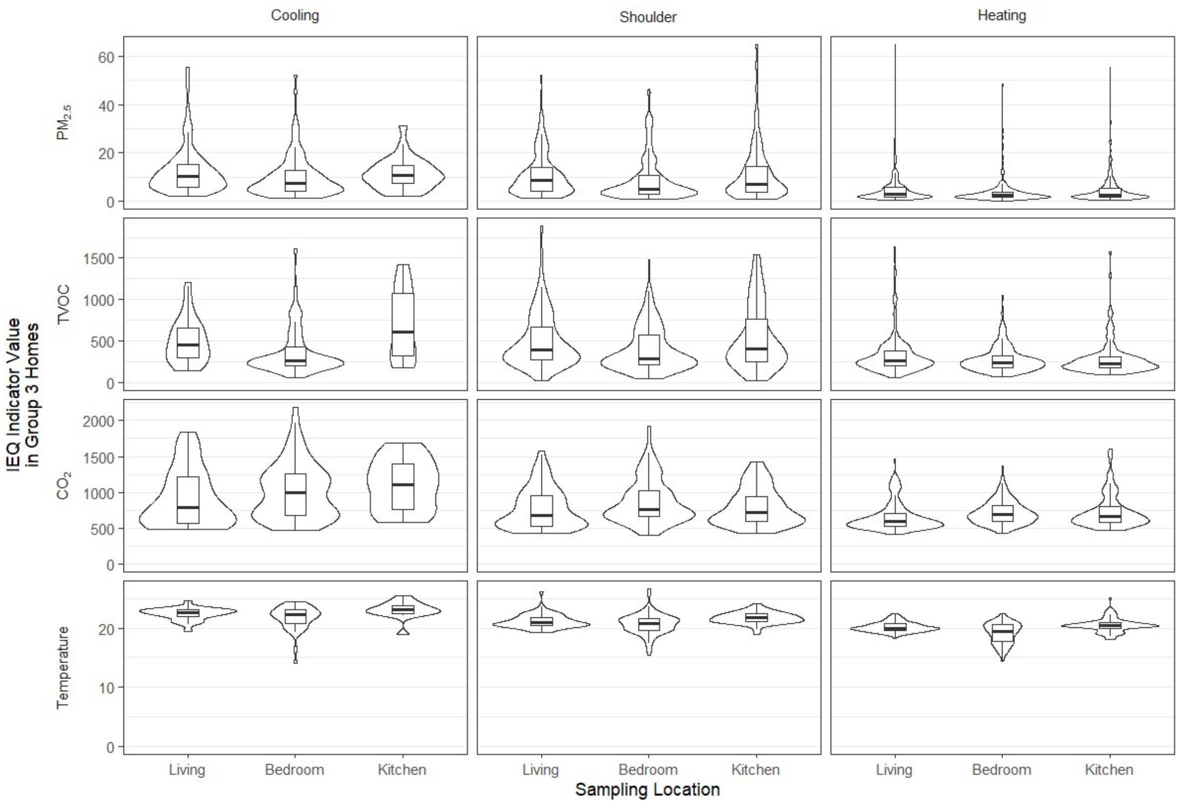


Figure 31: Distributions of time-averaged hourly IEQ parameter values from indoor rooms in Group 3 Homes (homes with definable cooling period), shown separately by season and sampling location. Values from all rooms in all Group 2 homes pooled. Units: $PM_{2.5}$ (µg/m^3), TVOC (ppb), CO_2 (ppm), Temperature (^{o}C).

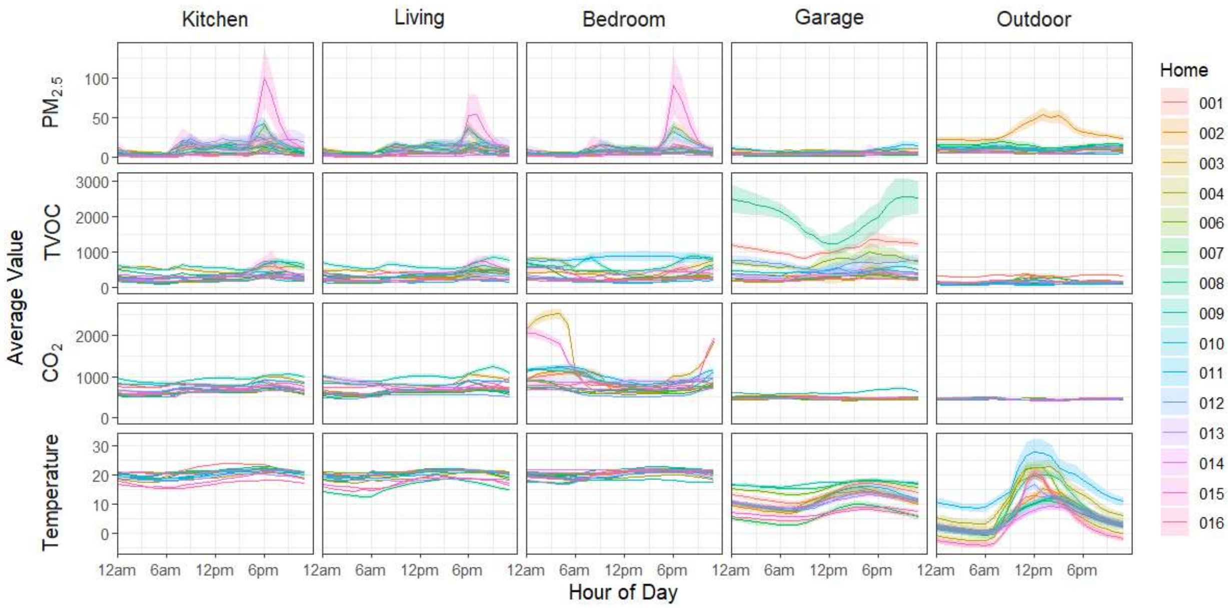


Figure 32: Average hour-of-day concentrations of $PM_{2.5}$ in µg/m³ (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in °C (bottom row) calculated from average hourly values over entire monitoring period (six to ten months) by home, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean.

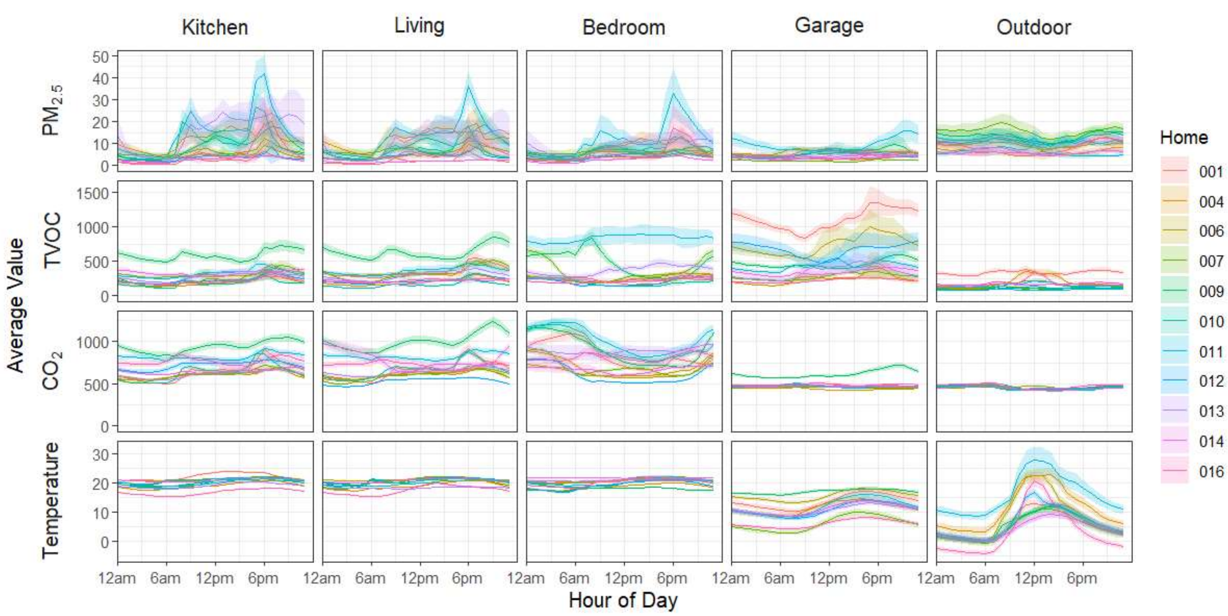


Figure 33: Average hour-of-day concentrations of $PM_{2.5}$ in µg/m³ (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in °C (bottom row) calculated from average hourly values over entire monitoring period (six to ten months) by home, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean. Sensors from Homes 2, 3, 8, and 15 omitted to give closer look at variation for most homes.

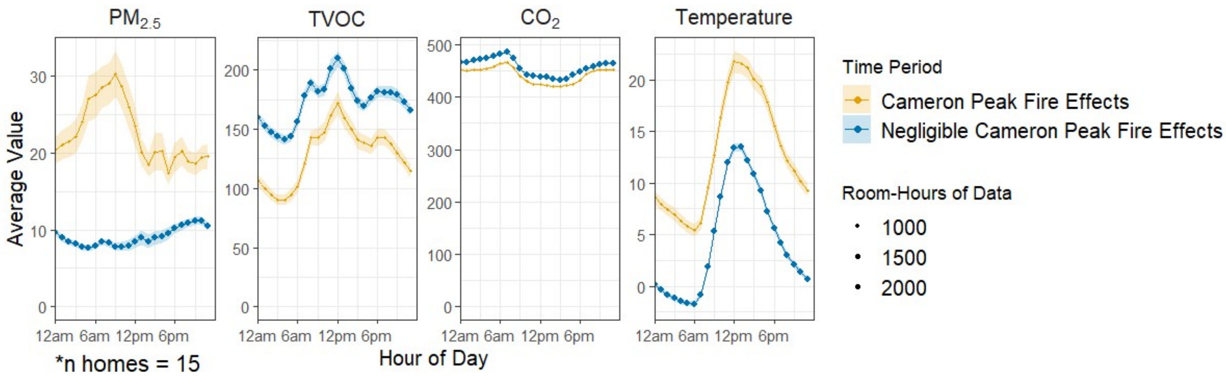


Figure 34: Average hour of day values of $PM_{2.5}$ in µg/m^3, TVOC in ppb, CO_2 in ppm, and temperature in ^{o}C calculated from average hourly values from kitchen, living room, and bedroom sensors in all homes pooled together, separated by time period during which data was collected. "Cameron Peak Fire Effects" time period (August 13th, 2020 to October 31st, 2020) is when the Cameron Peak wildfire was believed to have impacts on outdoor air quality; "Negligible Cameron Peak Fire Effects" time period (all other dates during the study period) is when impacts on air quality form the Cameron Peak wildfire were assumed negligible. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean. Size of point is proportional to the amount of hourly average values used to calculate an average value for each hour of day.

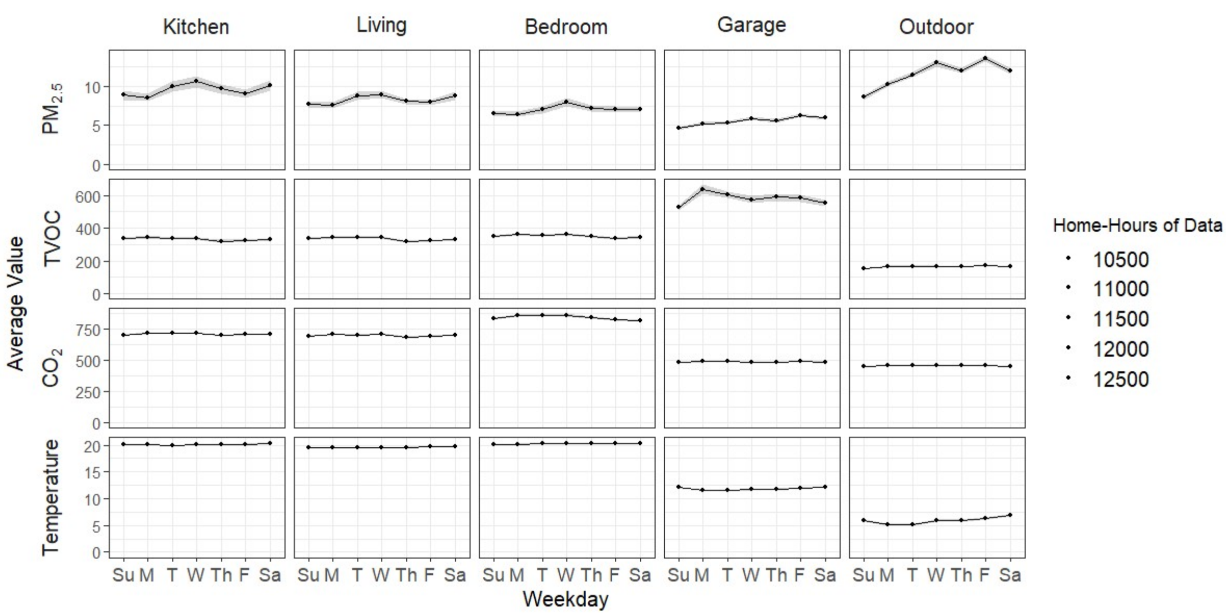


Figure 35: Average weekday concentrations of $PM_{2.5}$ in µg/m^3 (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in ^{o}C (bottom row) calculated from average hourly values over entire monitoring period for all homes pooled together, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean.

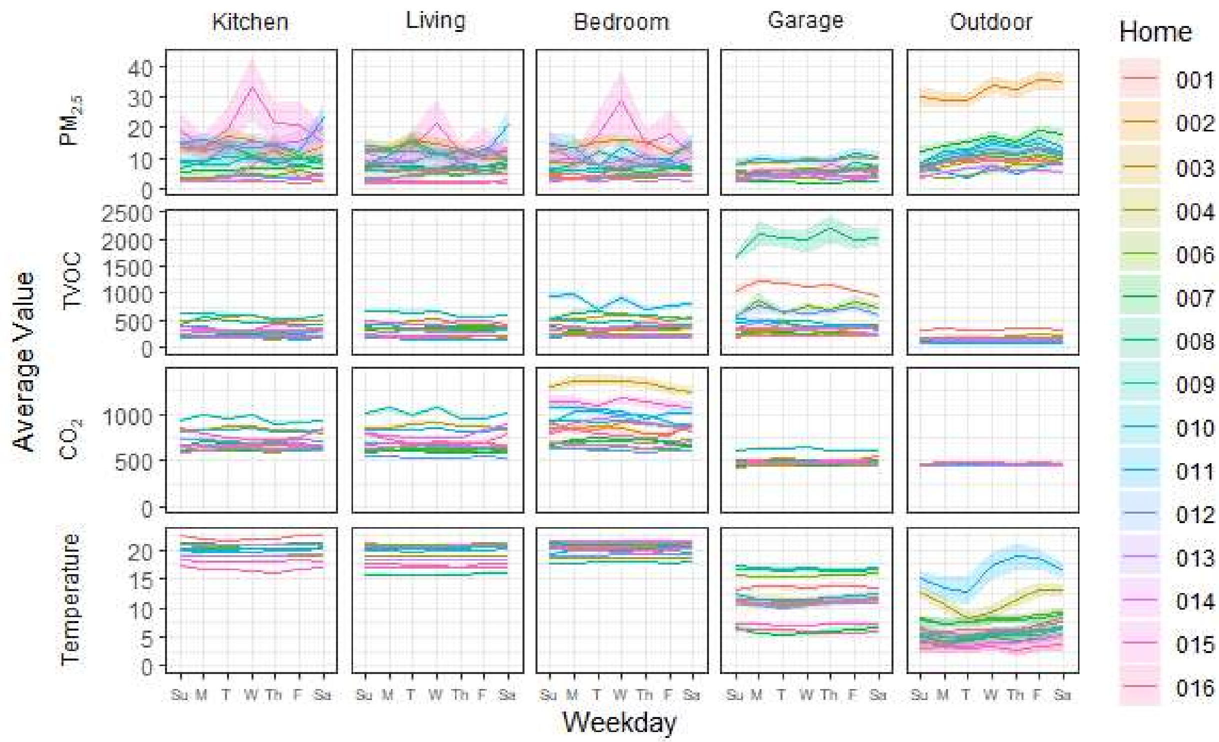


Figure 36: Average weekday concentrations of $PM_{2.5}$ in μg/m³ (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in °C (bottom row) calculated from average hourly values over entire monitoring period (six to ten months) by home, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean.

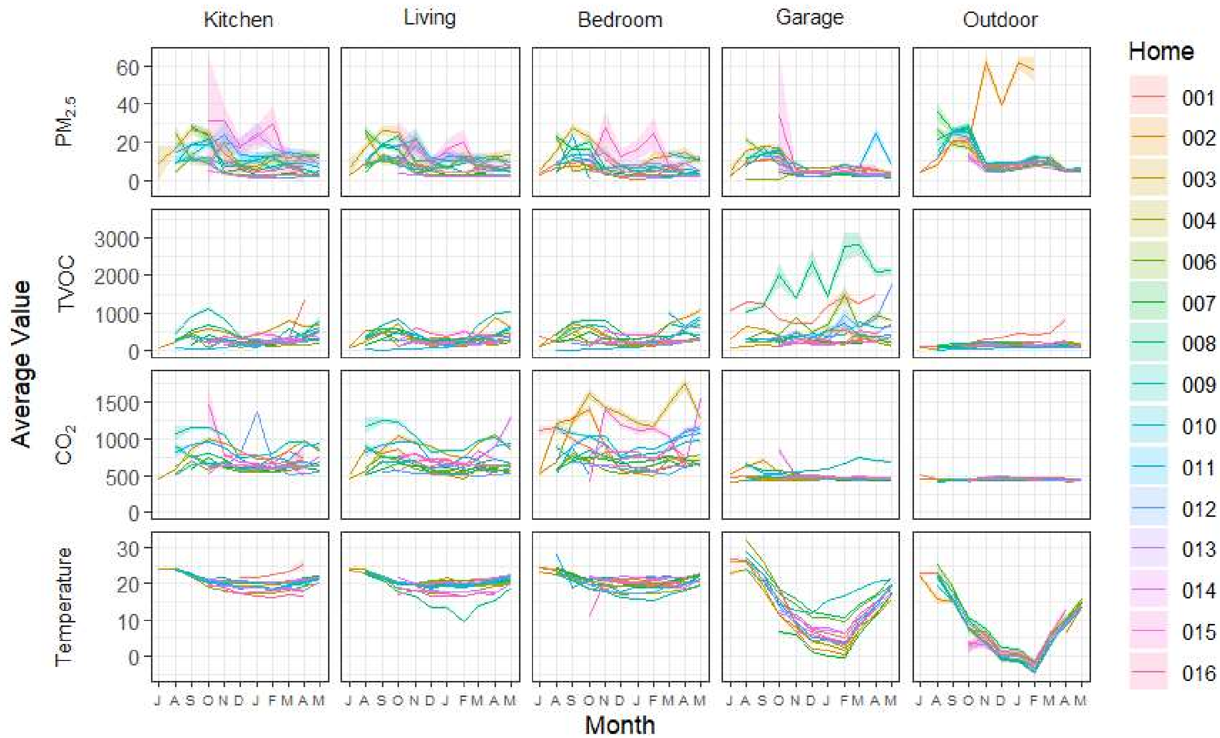


Figure 37: Average monthly concentrations of $PM_{2.5}$ in µg/m³ (top row), TVOC in ppb (second row), and CO_2 in ppm (third row), and temperature in °C (bottom row) calculated from average hourly values over entire monitoring period by home, disaggregated by monitor location. Line plot represents hour of day arithmetic mean values, and transparent ribbon represents limits of the 95% confidence interval around the mean.

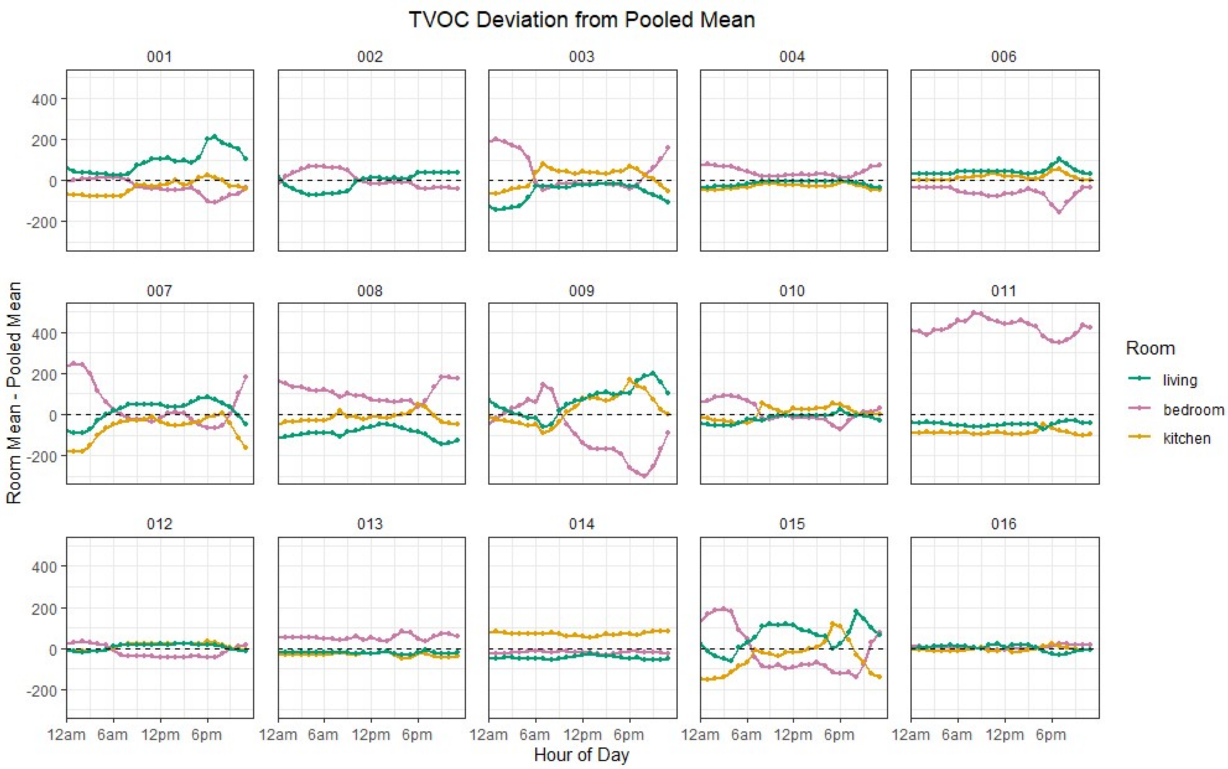


Figure 38: Pooled (between kitchen, living room, and bedroom) hour of day TVOC arithmetic mean subtracted from the hour of day TVOC arithmetic mean for each room at each hour of day over the entire monitoring period, plotted separately for all homes. Units in ppb. Confidence intervals were omitted to increase clarity.

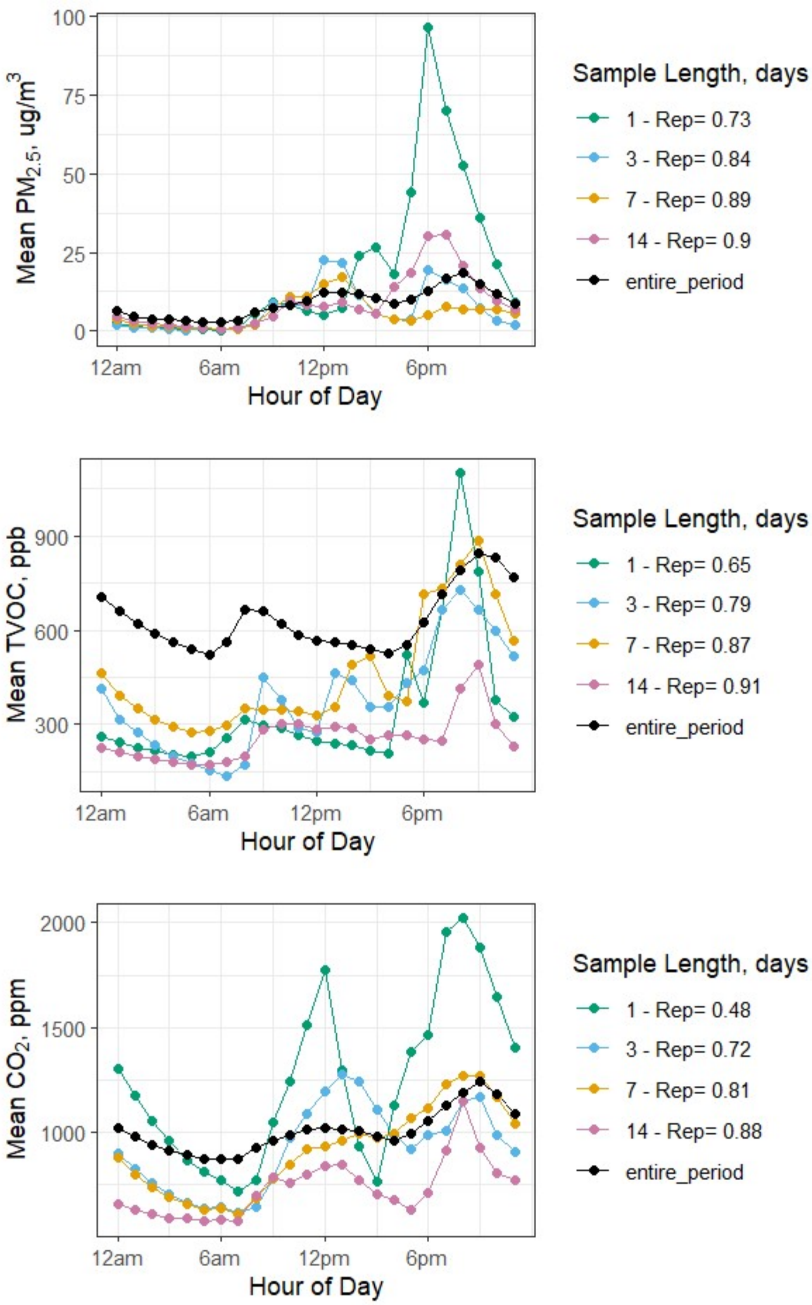


Figure 39: Hour-of-day arithmetic mean values calculated from samples of multiple lengths compared to those calculated from the entire monitoring period from the same condition for $PM_{2.5}$ (a, top), TVOCs (b, middle), and CO_2 (c, bottom). The considered condition was Home 9 living room in the heating season. Time-structured temporal representativeness (Rep) values for each sample are noted in the legend. Confidence intervals around mean values were omitted for clarity.

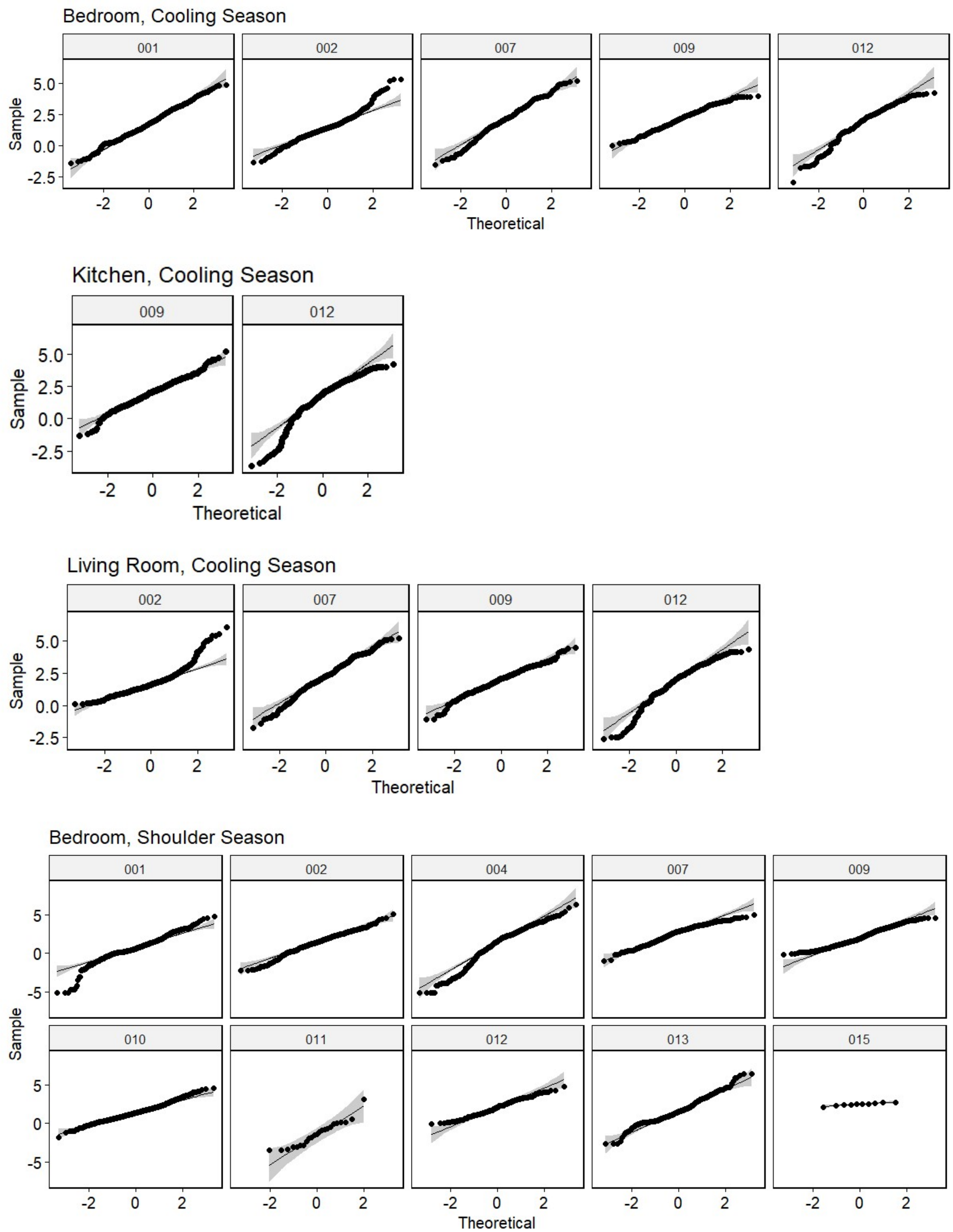


Figure 40: QQ Plots for $PM_{2.5}$ measurements for each home-room-season condition.

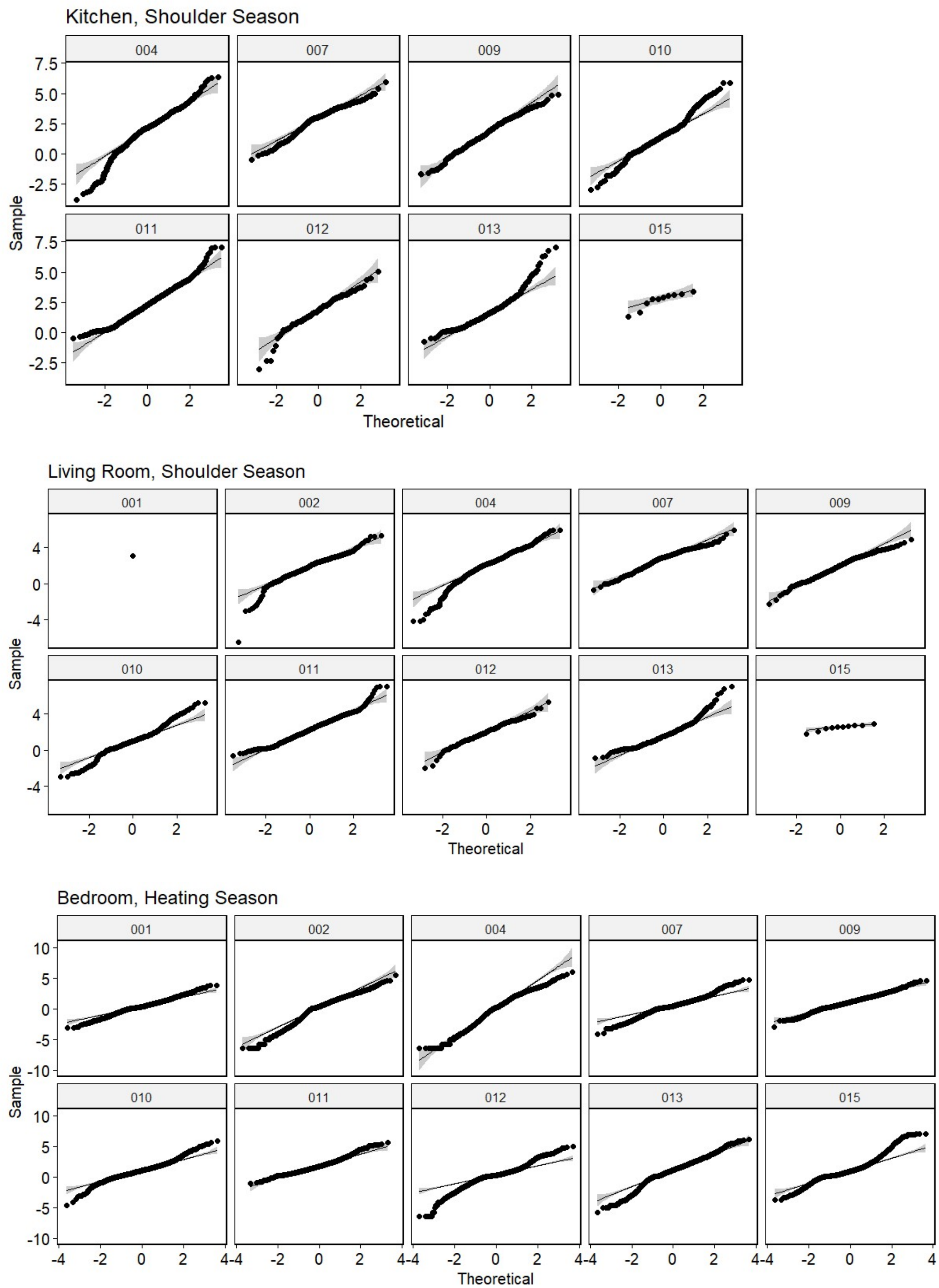
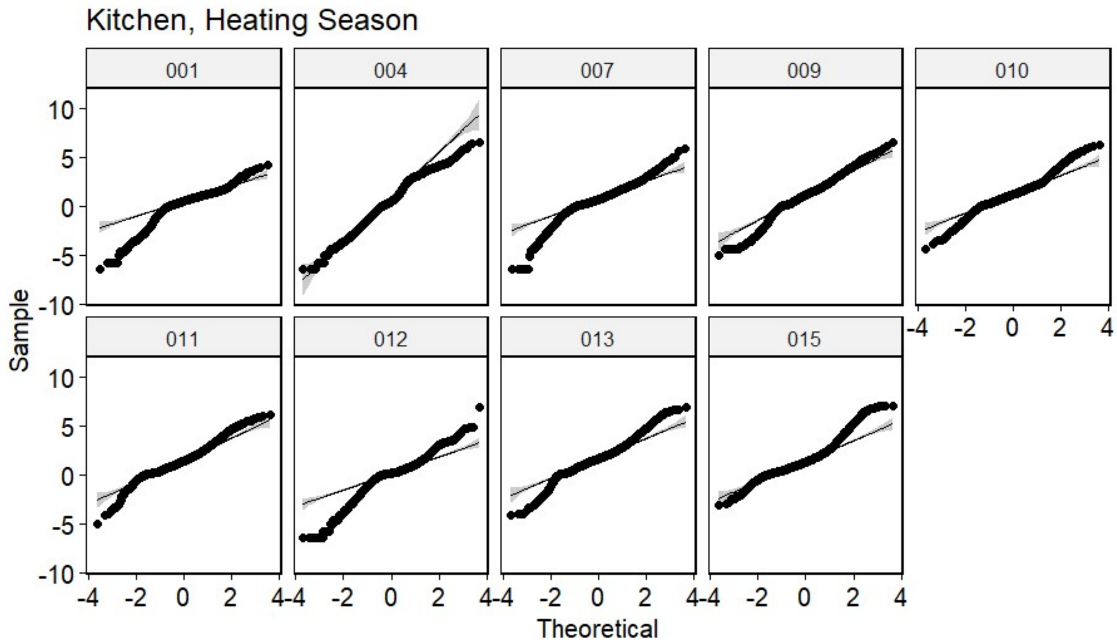


Figure 40 cont.: QQ Plots for $PM_{2.5}$ measurements for each home-room-season condition.



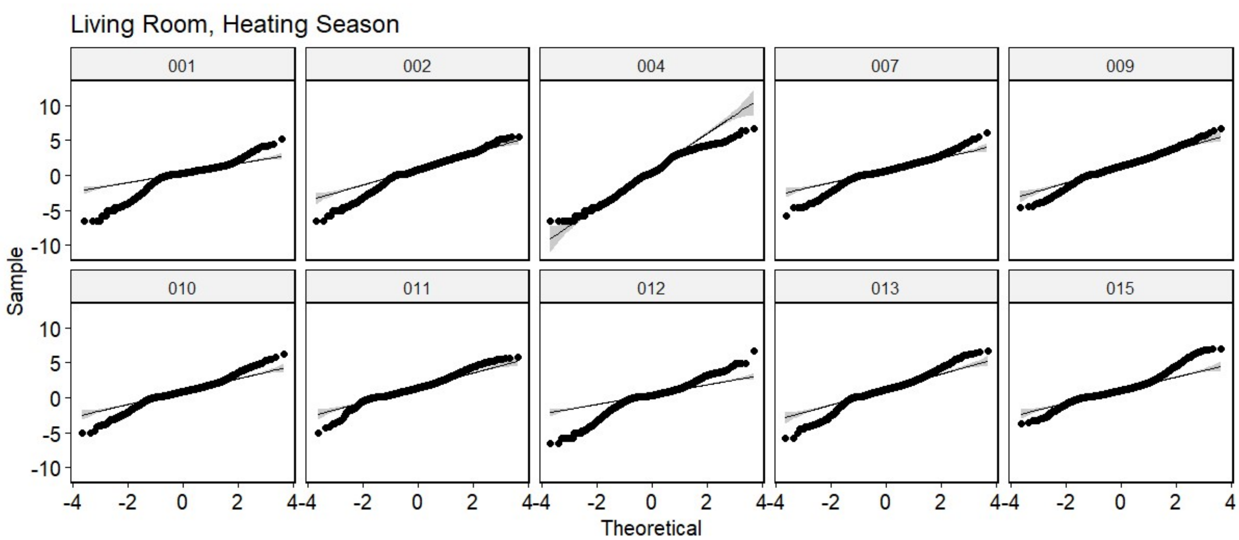


Figure 40 cont.: QQ Plots for $PM_{2.5}$ measurements for each home-room-season condition.

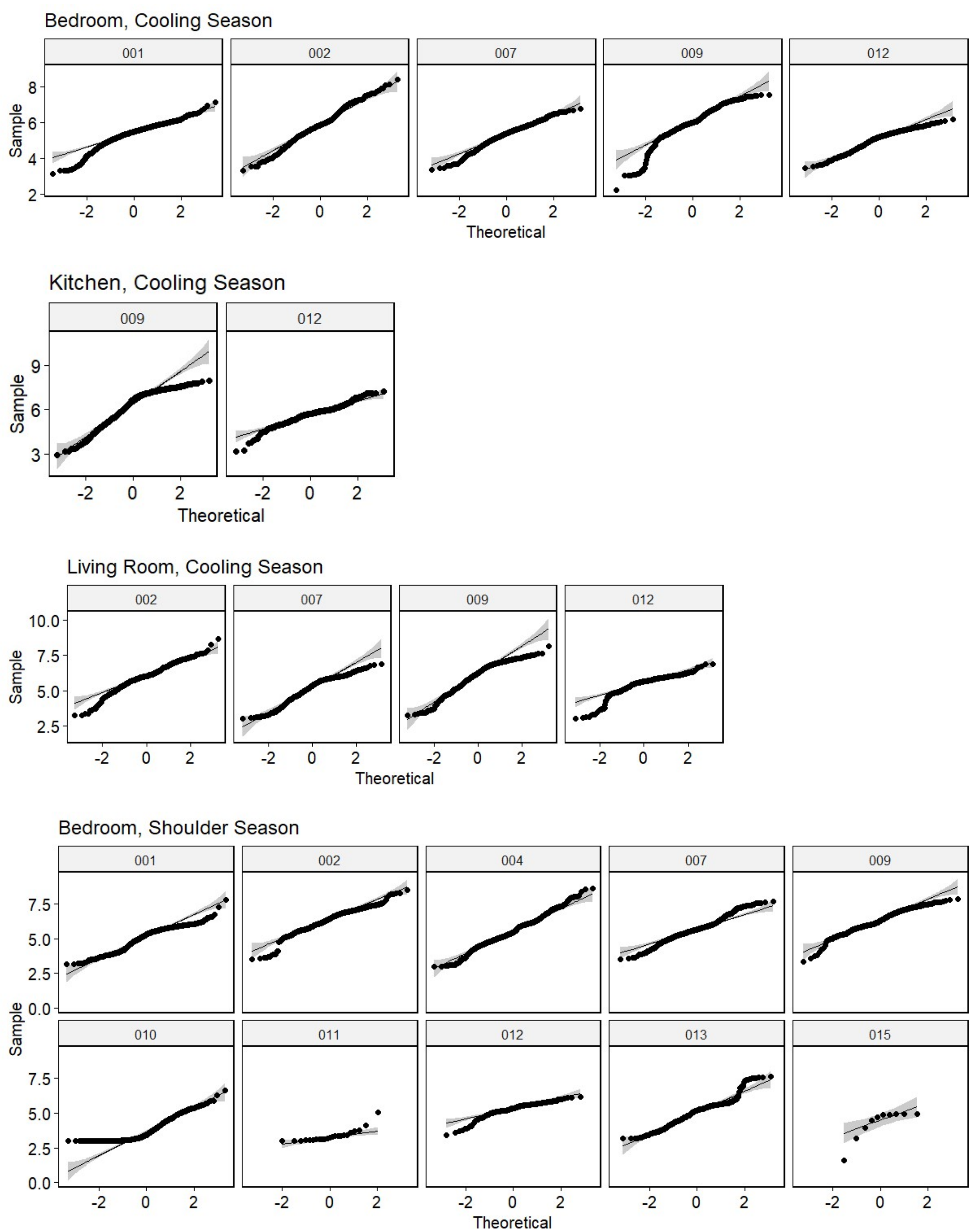
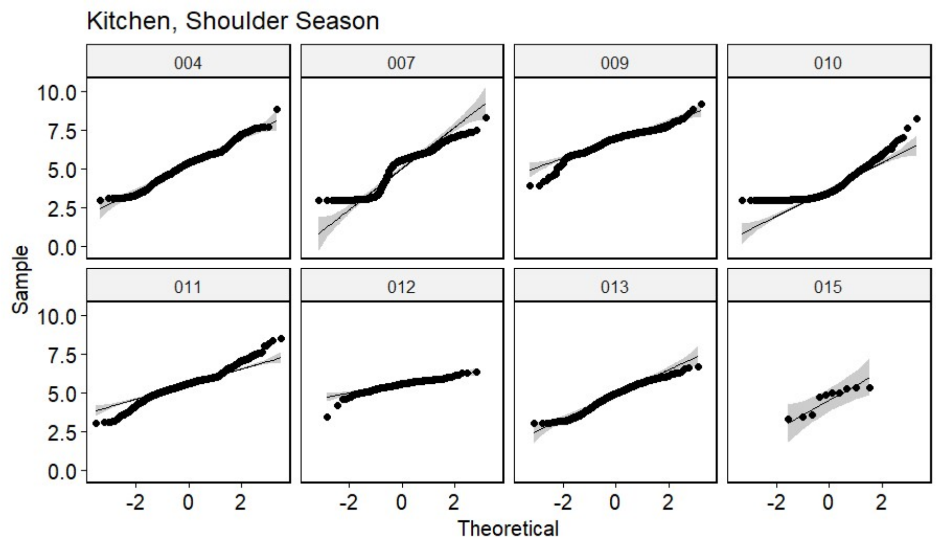
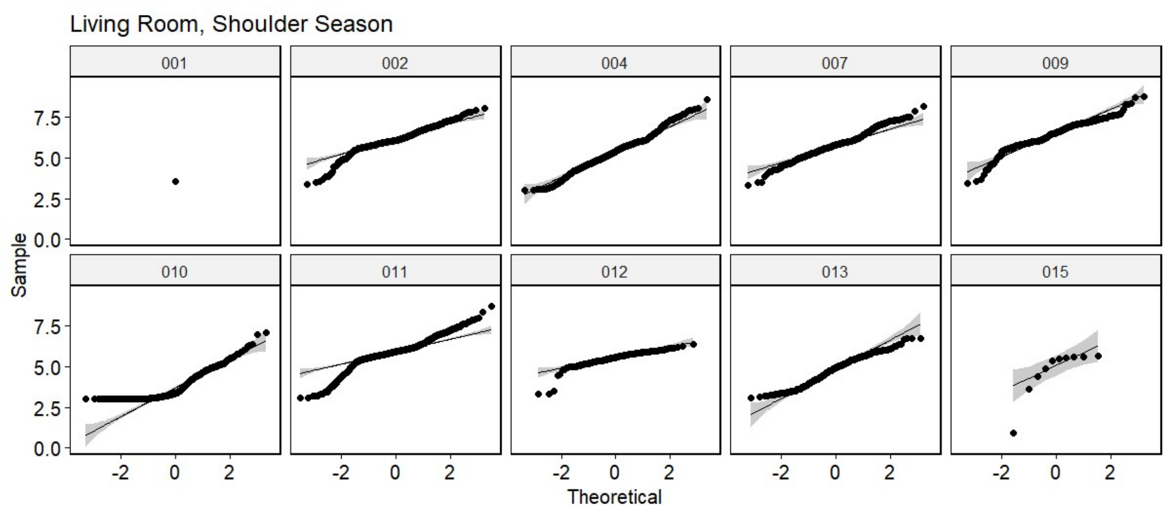


Figure 41: QQ Plots for TVOC measurements for each home-room-season condition.





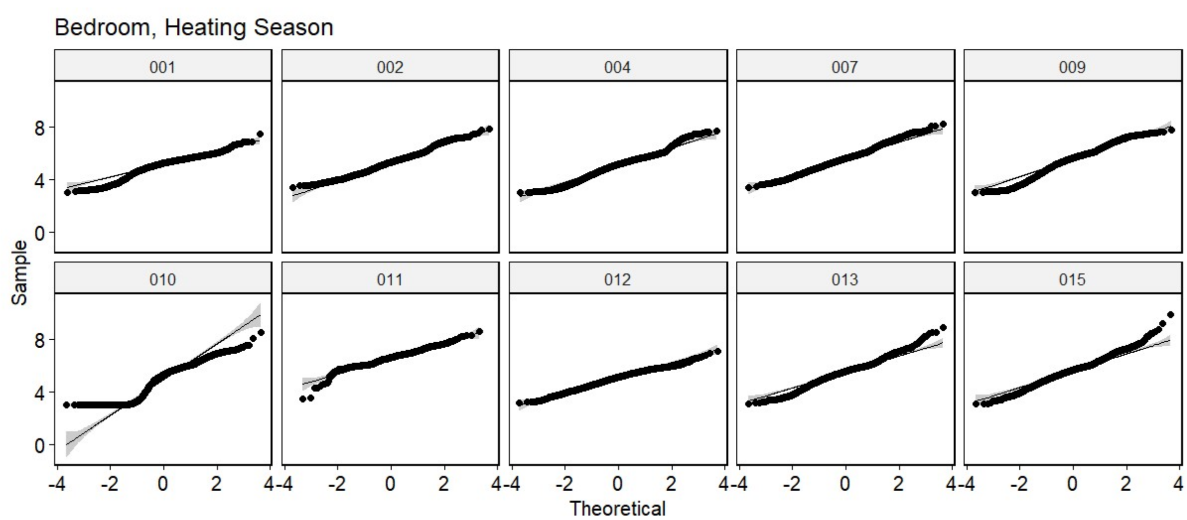
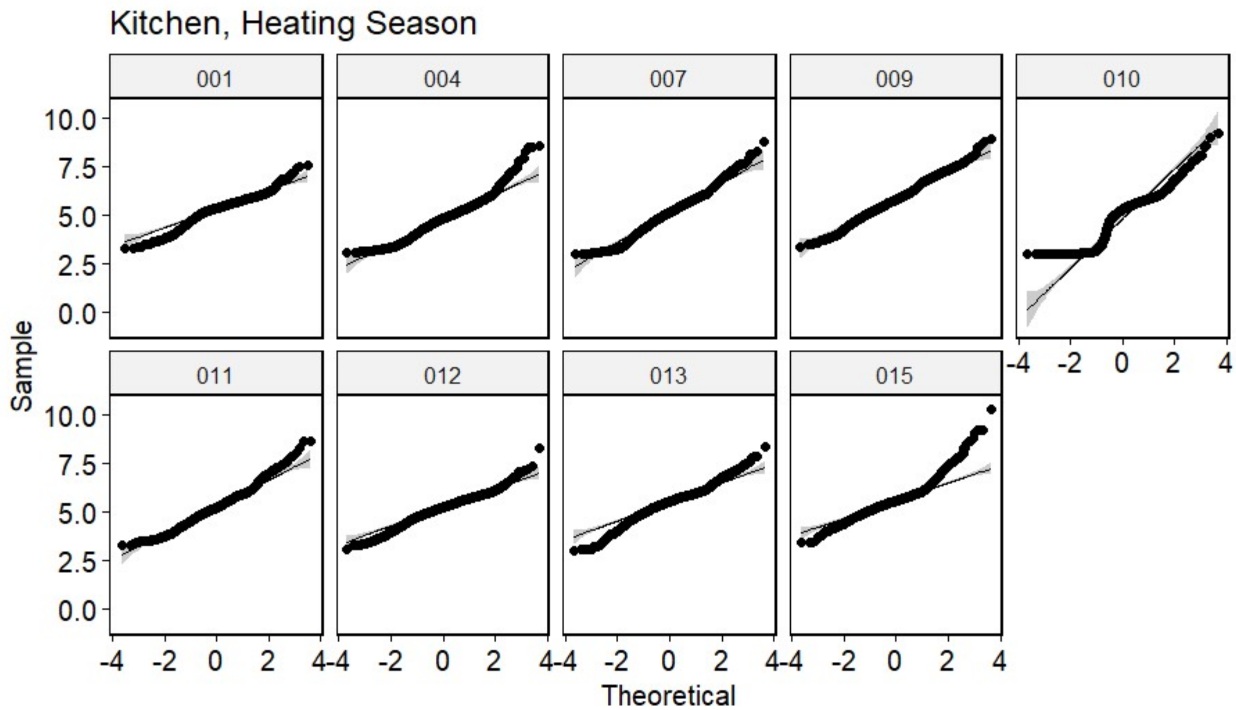


Figure 41 cont.: QQ Plots for TVOC measurements for each home-room-season condition.



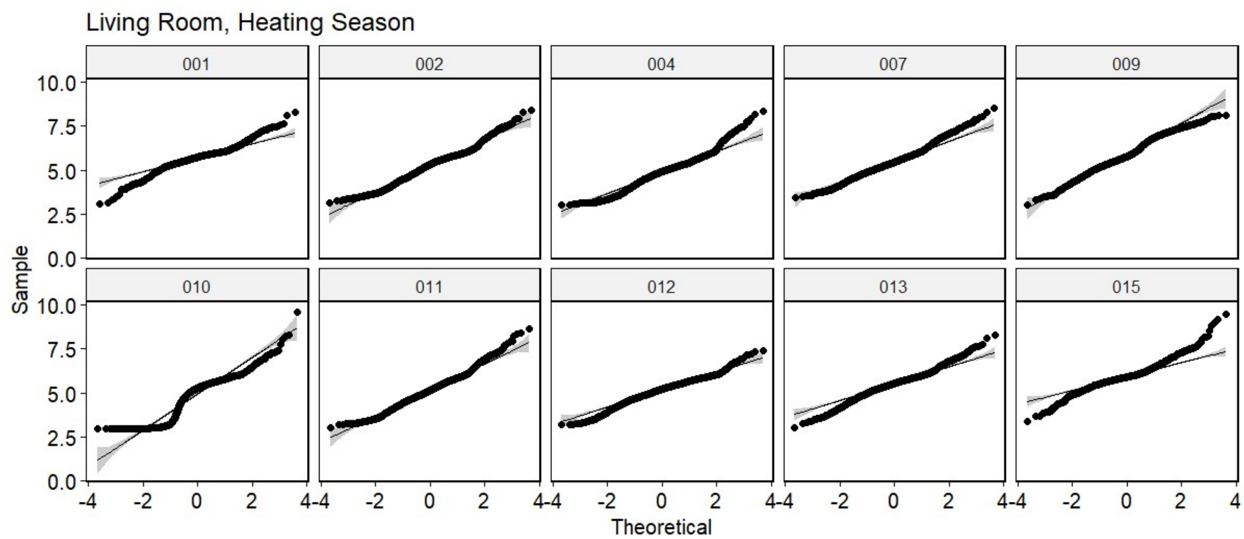


Figure 41 cont.: QQ Plots for TVOC measurements for each home-room-season condition.

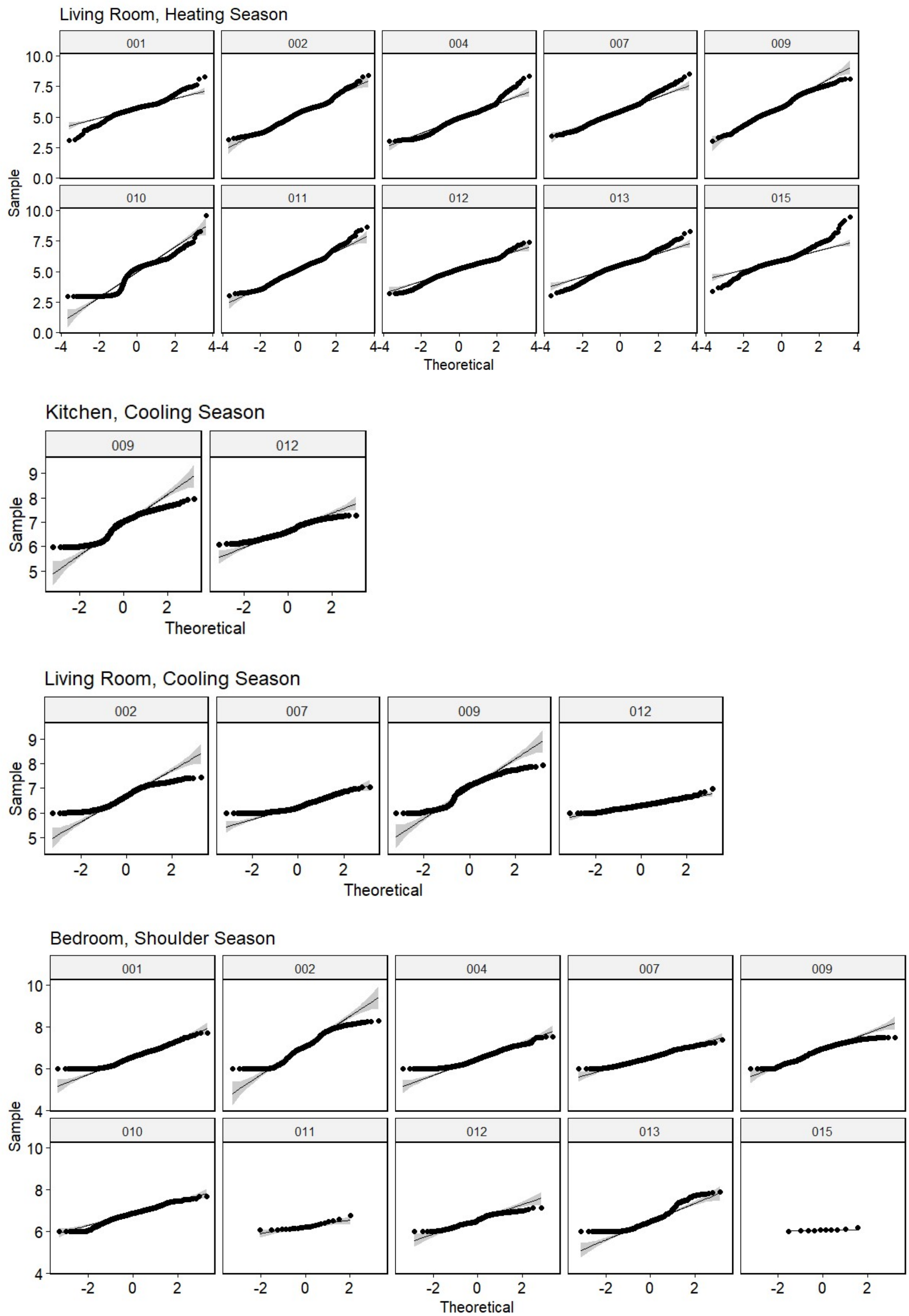


Figure 42: QQ Plots for CO_2 measurements for each home-room-season condition.

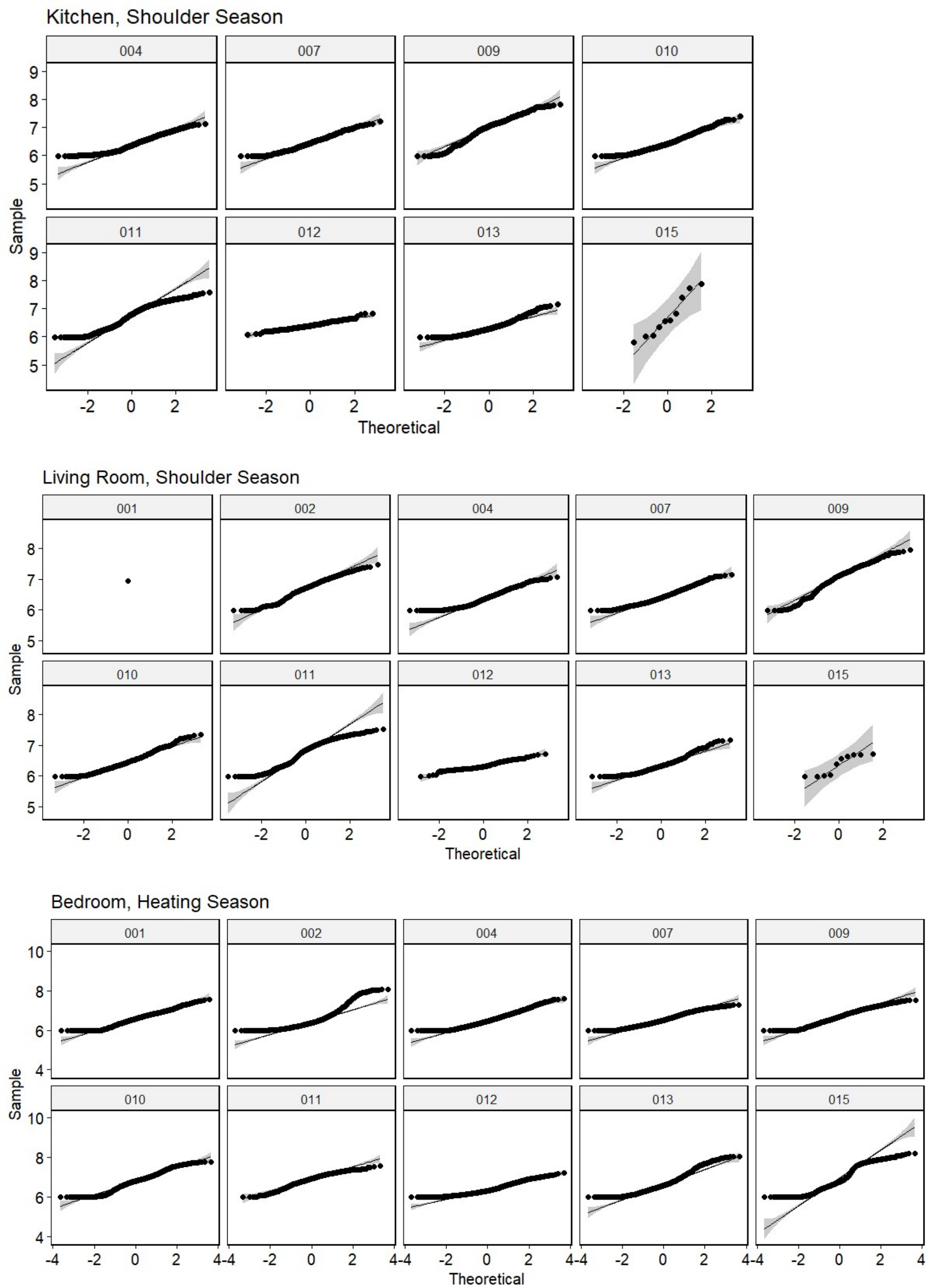


Figure 42 cont.: QQ Plots for CO_2 measurements for each home-room-season condition.

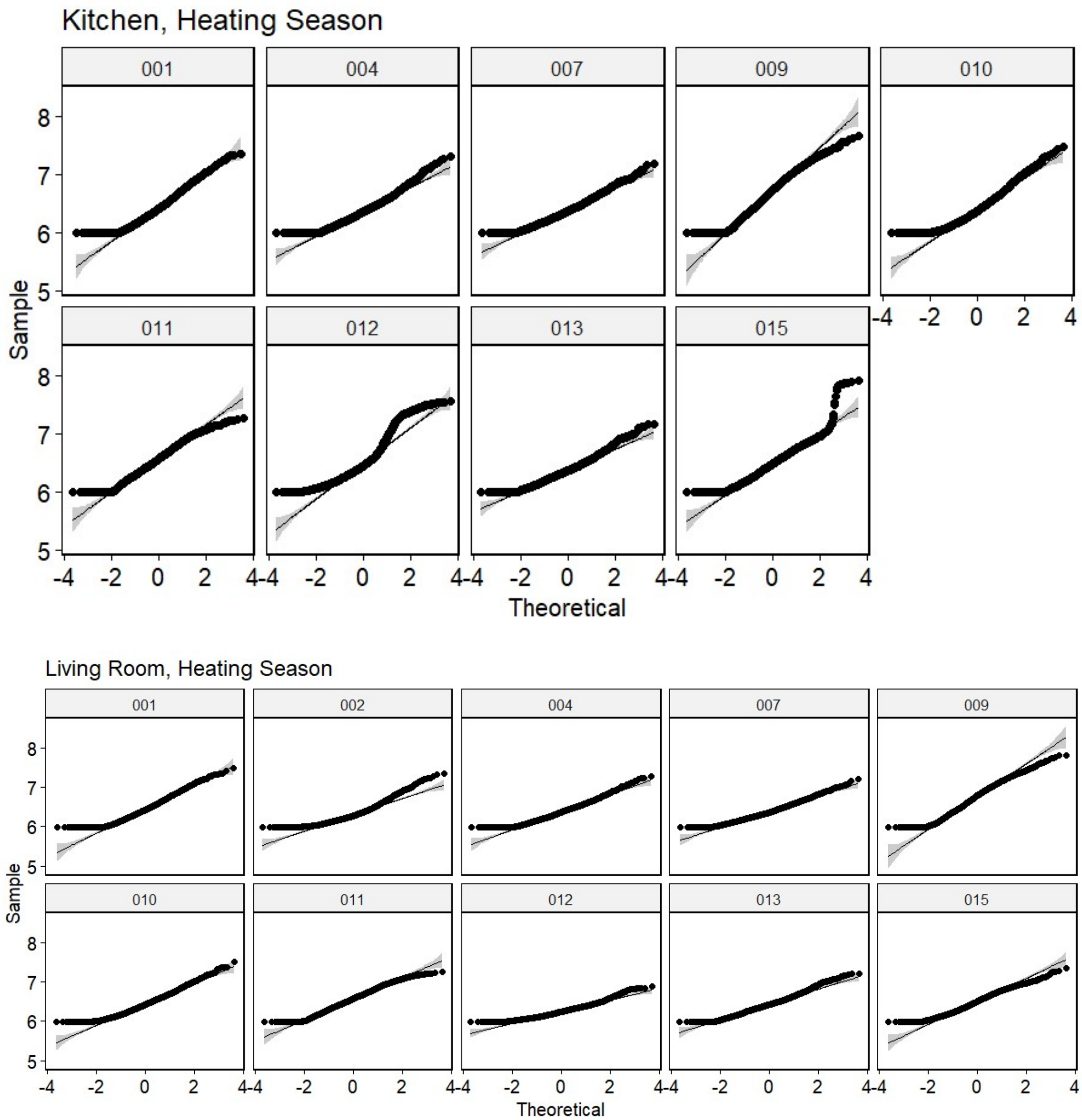


Figure 42 cont.: QQ Plots for CO_2 measurements for each home-room-season condition.

Table 12: Results from Akaike information criterion (AIC) calculations for three distribution types (log-normal, Weibull, and gamma) for three IAQ metrics ($PM_{2.5}$, TVOC, and CO_2) for all home-season-room conditions. Minimum AIC value implies the best fitting distribution type out of the three considered. "N Minimum Value" is the number of home-season-room conditions for which the given distribution type had the lowest AIC value for the given IAQ metric. "% Minimum Value" is the percentage of home-season-room conditions for which the given distribution type had the lowest AIC value for the given AIQ metric.

Metric	Distribution Type	N Minimum Value	% Minimum Value
$PM_{2.5}$	Log-Normal	66	76%
	Weibull	21	24%
	Gamma	0	0%
TVOC	Log-Normal	66	76%
	Weibull	0	0%
	Gamma	21	24%
CO_2	Log-Normal	84	97%
	Weibull	3	3%
	Gamma	0	0%

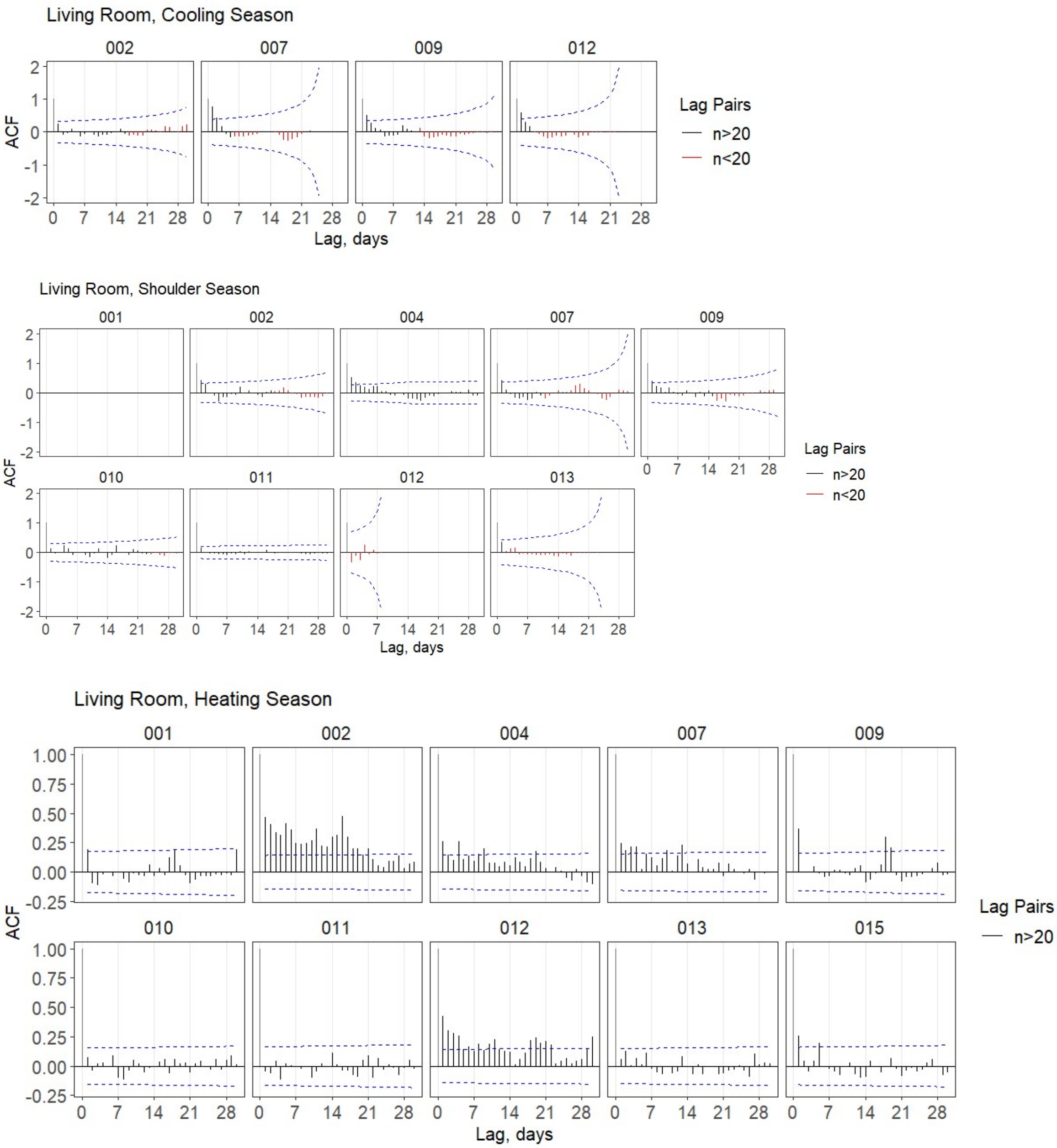


Figure 43: Correlograms of $PM_{2.5}$ data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

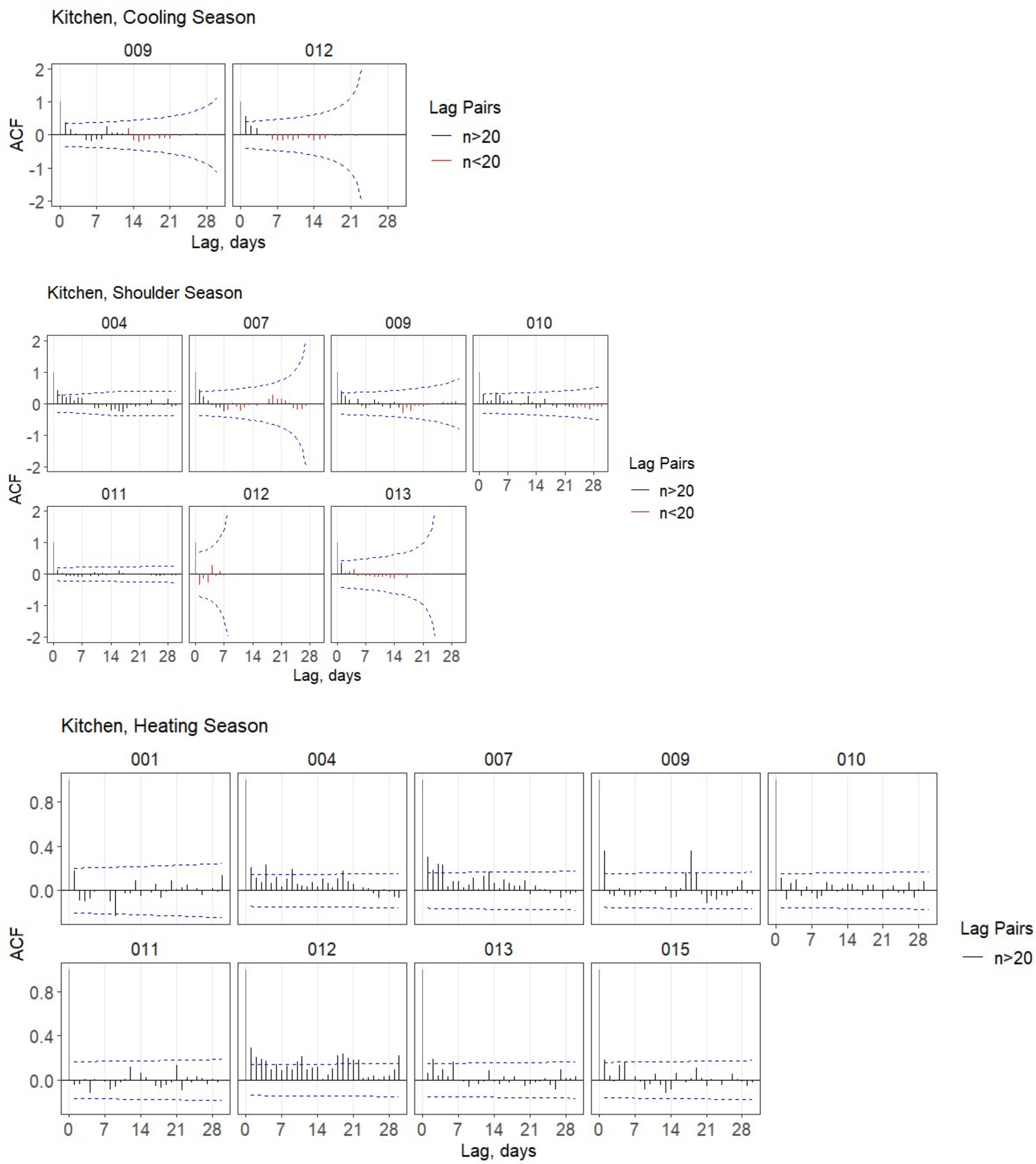


Figure 43 cont.: Correlograms of $PM_{2.5}$ kitchen data for all homes (number of home on top of plot) in cooling (top) shoulder (middle) and heating (bottom) seasons. Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

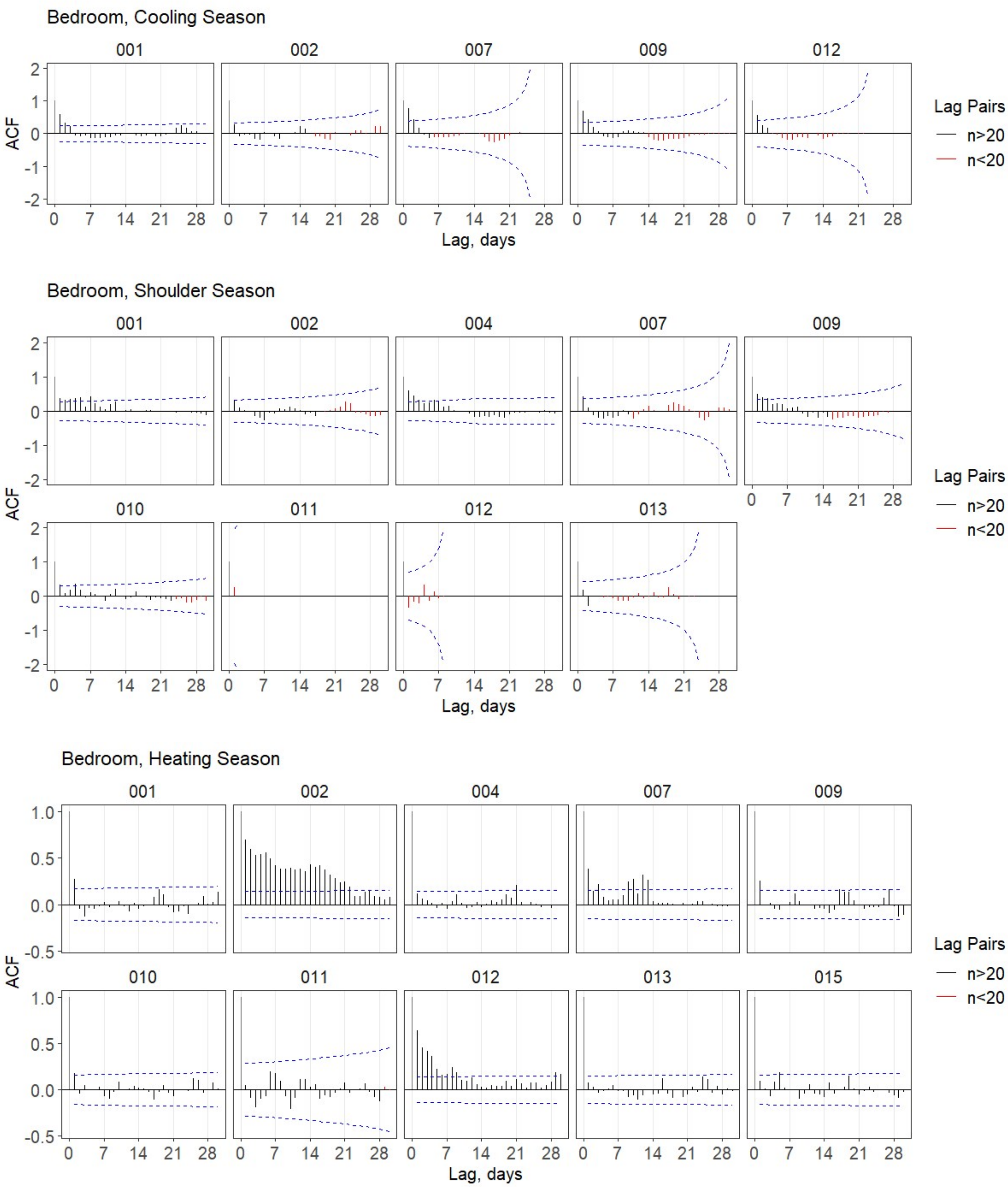


Figure 43 cont.: Correlograms of $PM_{2.5}$ kitchen data for all homes (number of home on top of plot) in cooling (top) shoulder (middle) and heating (bottom) seasons. Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

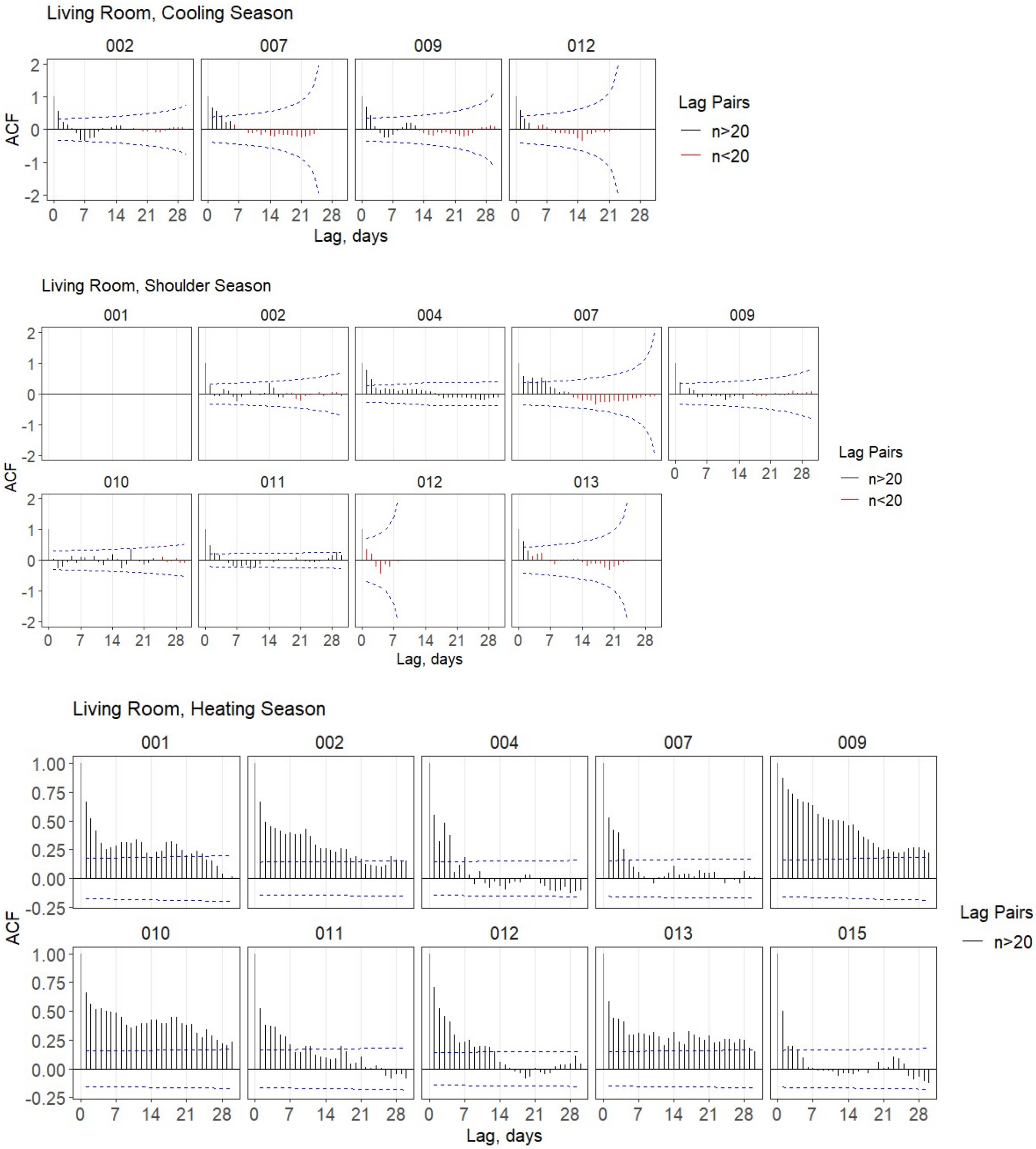


Figure 44: Correlograms of TVOC data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

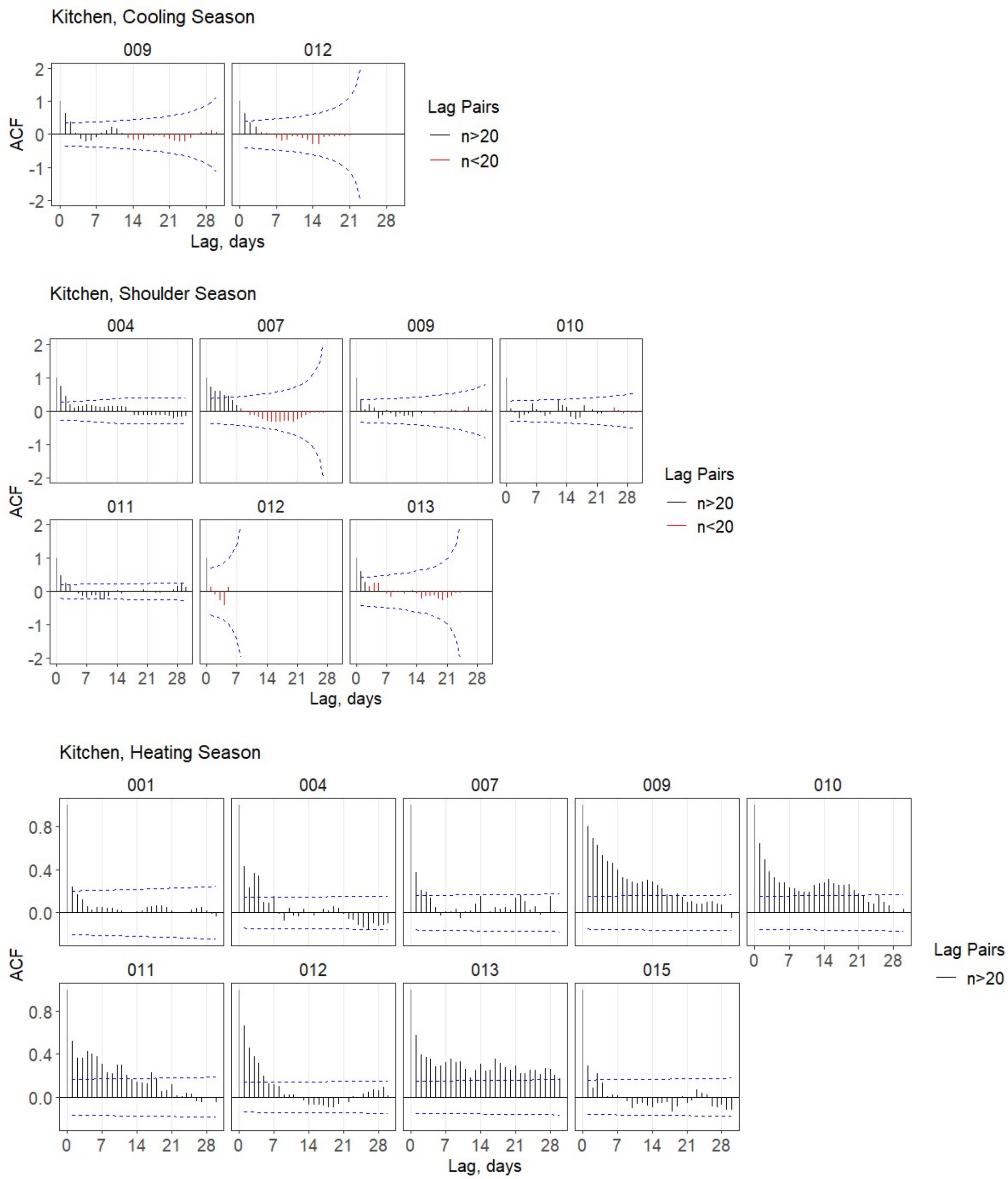


Figure 44 cont.: Correlograms of TVOC data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

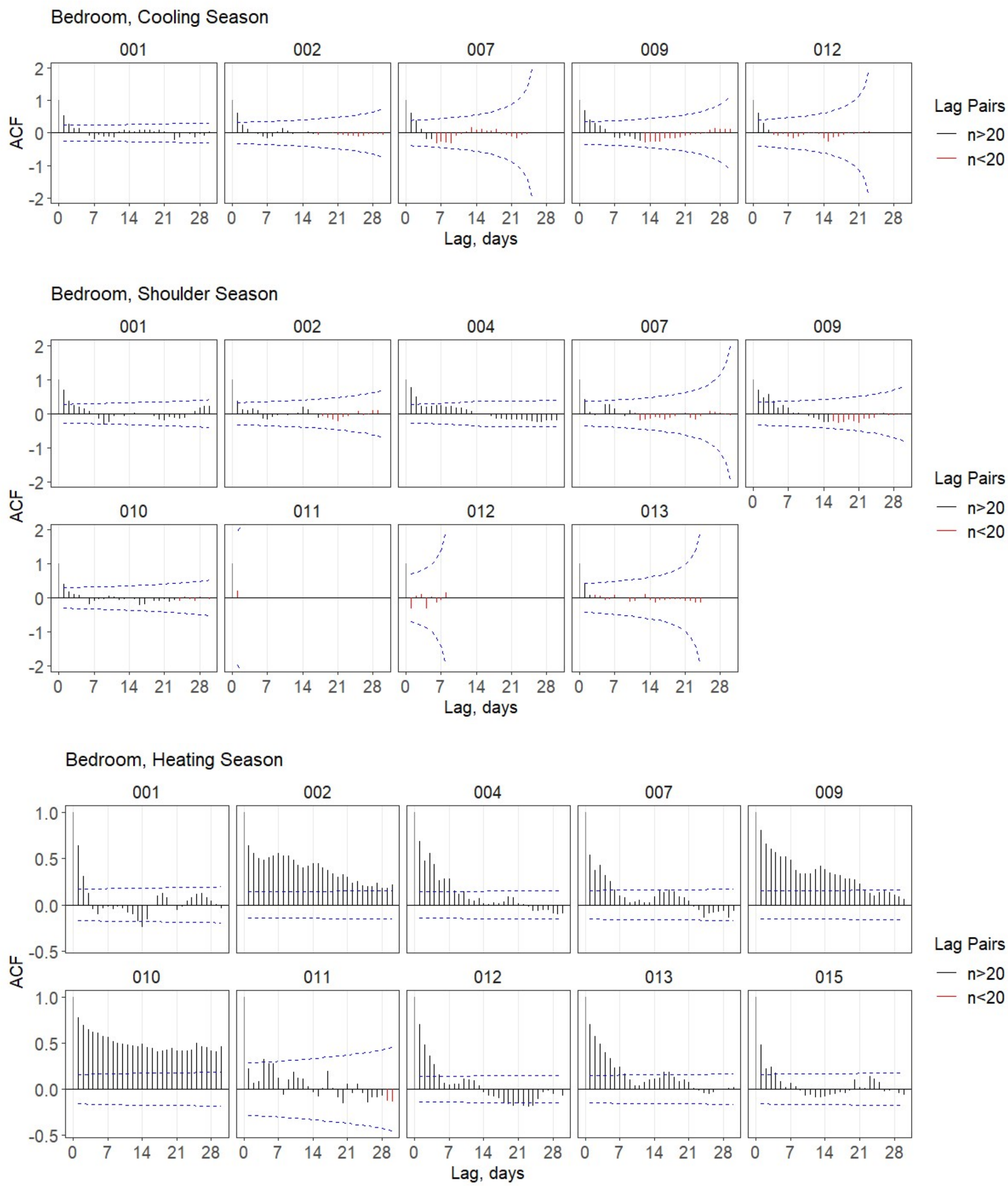


Figure 44 cont.: Correlograms of TVOC data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

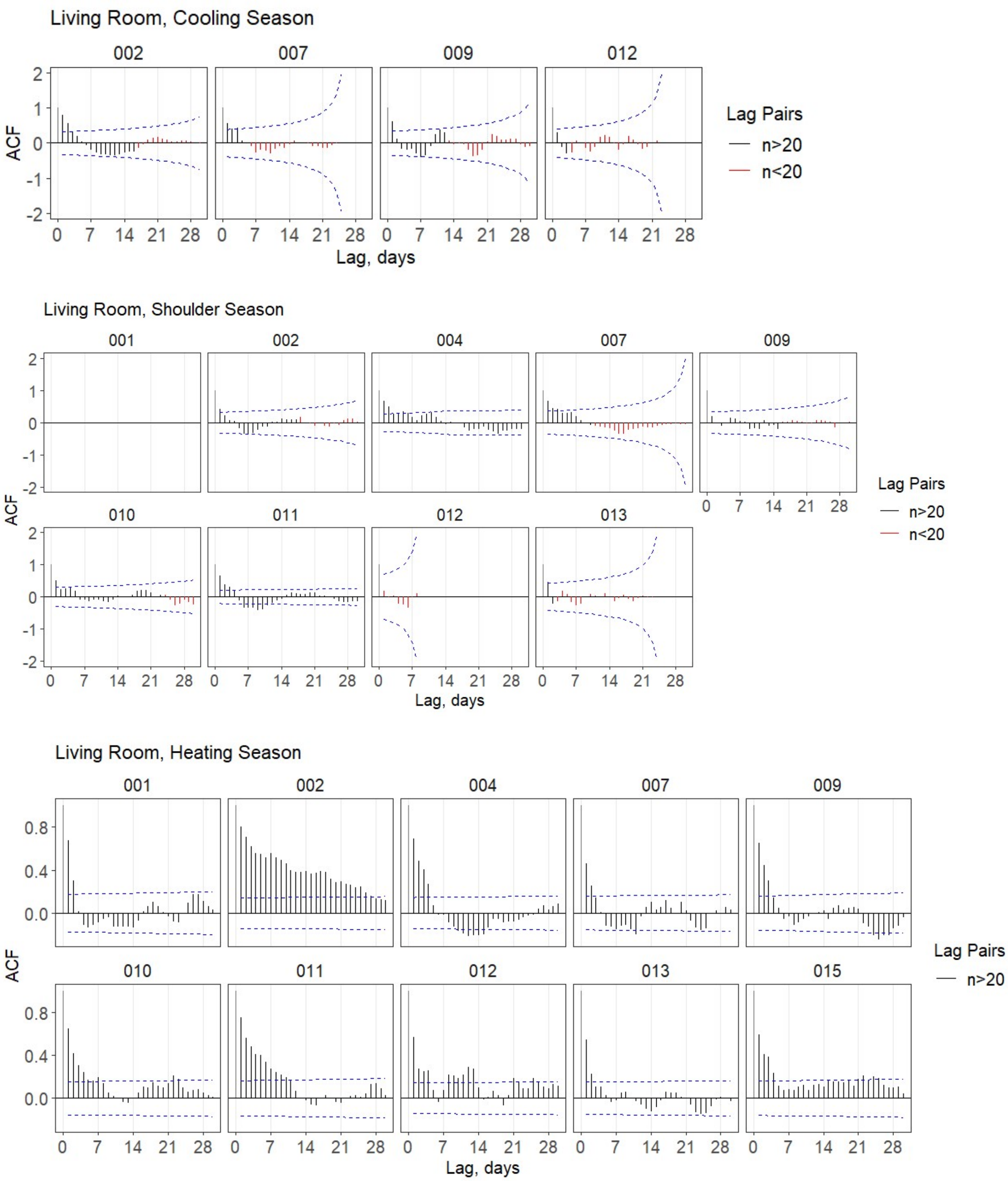


Figure 45: Correlograms of CO_2 data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

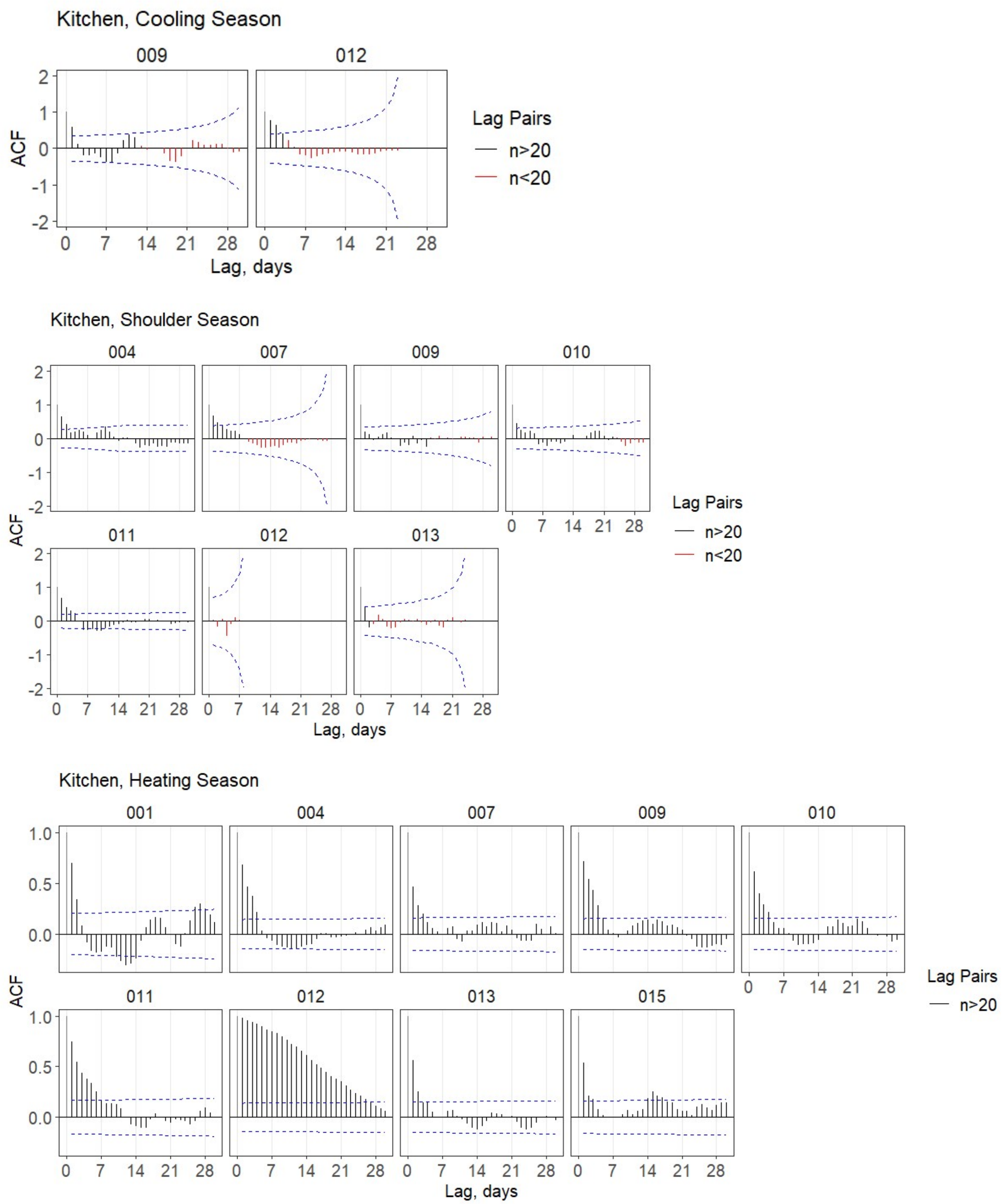


Figure 45 cont.: Correlograms of CO_2 data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

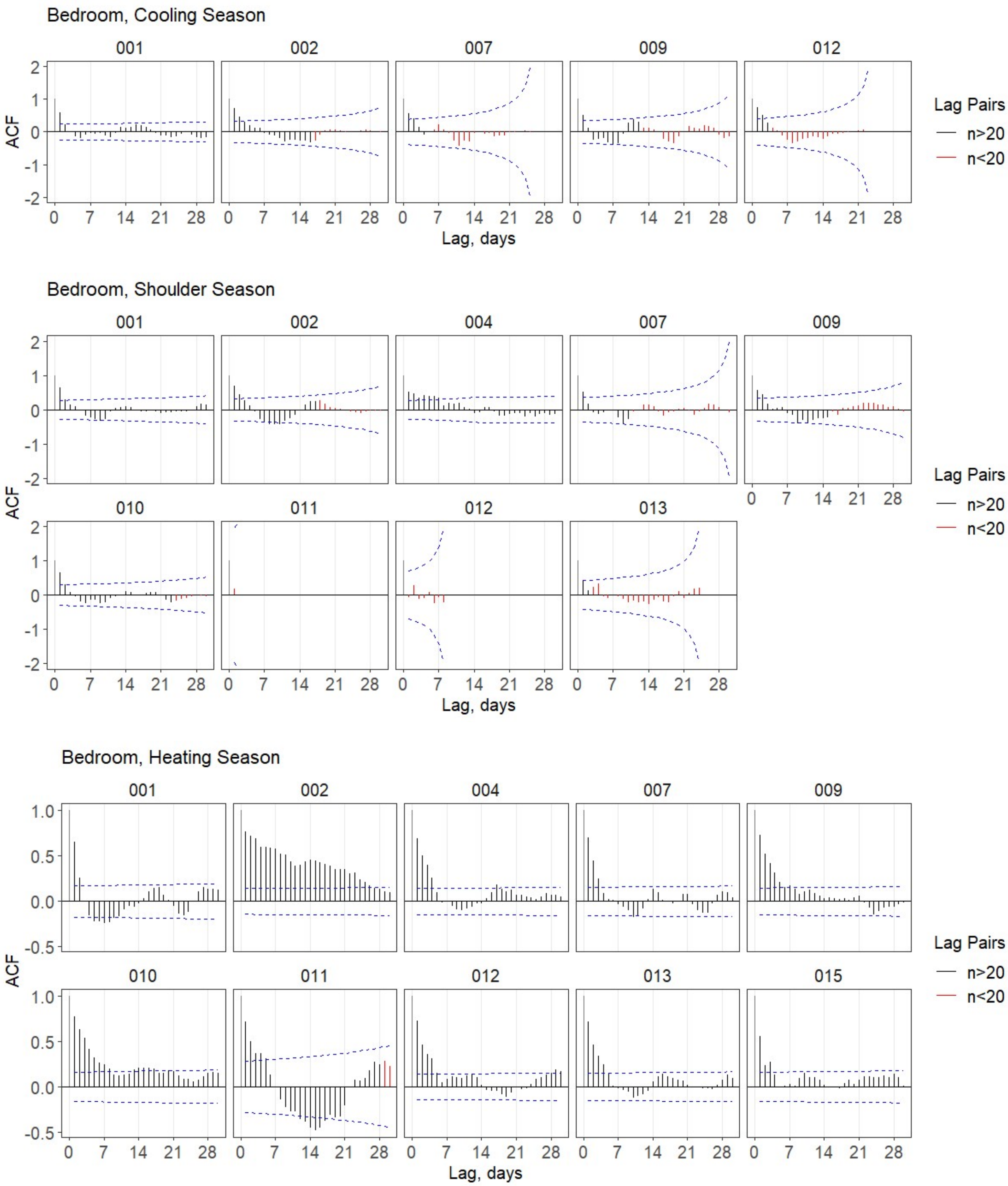


Figure 45 cont.: Correlograms of CO_2 data for all homes (number of home on top of individual plot) grouped by room and season (noted above plot groups). Black line segments imply more than 20 pairs of daily data pairs were available to calculate the autocorrelation factor (ACF) value at a given lag, while red values imply less than 20 data pairs. Dashed blue lines represent 95% confidence intervals. ACF values that are greater than the high confidence interval or less than the low confidence interval for the given lag number are considered significant.

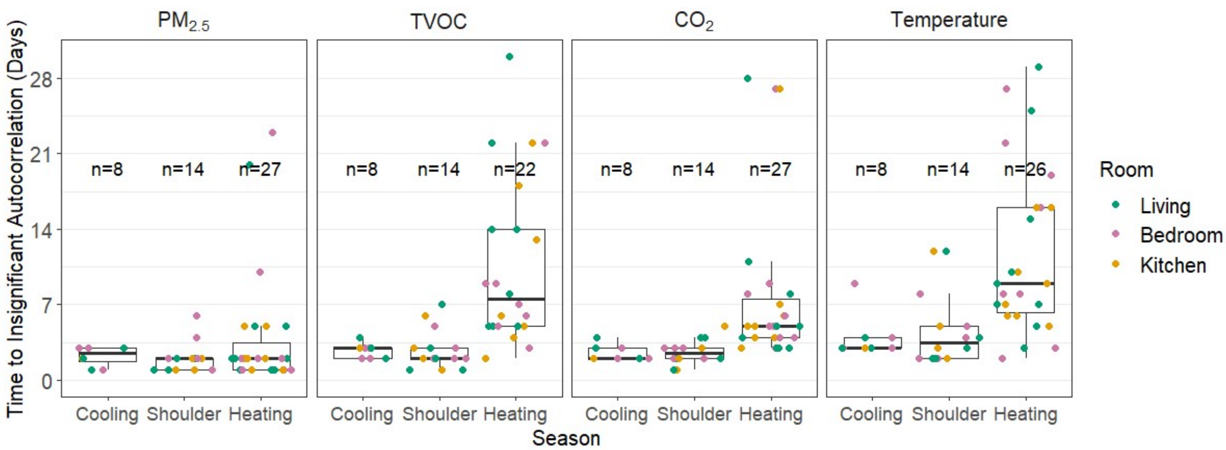


Figure 46: Number of days (in lagged correlations) before insignificant autocorrelation is reached for each IEQ indicator, homes pooled, stratified by season and colored by room. Room-season samples consisting of less than 25 days of data, missing more than 11.1% (1/9) of monitored days, or that did not reach insignificance prior to 30 days were omitted from analysis.