American Journal of Applied Sciences

Original Research Paper

Measurement of Cooling Load-Based Occupant Behavior in Under-Actuated Zones: A Time-Variance Approach

[1]Yaddarabullah, [1]Yusuf Maulana Akbar, [2]Aedah Binti Abd Rahman and [3]Amna Saad

[1]*Department of Informatics, Universitas Trilogi, Jakarta, Indonesia*
[2]*School of Science and Technology, Asia e University, Selangor, Malaysia*
[3]*Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia*

Article history
Received: 03-08-2023
Revised: 08-10-2023
Accepted: 26-10-2023

Corresponding Author:
Yaddarabullah
Department of Informatics,
Universitas Trilogi, Jakarta,
Indonesia
Email: yaddarabullah@trilogi.ac.id

Abstract: This research is dedicated to the enhancement of Heating, Ventilation, and Air-Conditioning (HVAC) systems through an innovative investigation of occupant behavior within under-actuated zones. These zones are strategically positioned to elevate indoor air quality and comfort through the intelligent utilization of HVAC systems. However, accurately measuring the cooling loads within these zones, which are inherently influenced by the presence of occupants, their activities, and appliance usage, presents a formidable challenge. To effectively tackle this challenge, our study introduces a pioneering methodology that intricately links cooling load measurements to occupant behavior in under-actuated zones. Central to this methodology is the comprehensive consideration of critical factors, including occupant count, activity patterns, and operation of diverse appliances. Notably, our analysis employs a dynamic temporal lens to scrutinize time variances and intervals. To discern underlying behavioral patterns in under-actuated zones, we employed the Hartigan's dip test, an analytical tool that captures the multimodal distribution of time variances. This highlights the intricate behavioral trends that are pivotal for optimizing HVAC strategies. Moreover, our analysis integrates advanced statistical techniques, such as polynomial regression, to evaluate cooling loads and determine the optimal time intervals. Through rigorous examination, we identified four zones exhibiting bimodal distributions and one zone characterized by a unique trimodal distribution. These findings not only establish appropriate time intervals for fine-tuning cooling loads based on occupant behavior but also yield robust correlations, with R-squared values exceeding 0.9. An essential facet of our study pertains to ethical dimensions. We meticulously addressed considerations concerning human subjects, ensuring unwavering integrity and adherence to the established ethical guidelines.

Keywords: Time Variance Analysis, Internal Cooling Load, HVAC System, Occupant Behavior, Under-Actuated Zone

Introduction

Heating, Ventilation and Air Conditioning (HVAC) systems, occasionally referred to as centralized air conditioning, are essential for controlling the temperature and humidity of the air in a room (Bahramnia *et al*., 2019). These systems function by adjusting the temperature to meet specific set point requirements, ensuring a comfortable environment for the occupants (Papadopoulos *et al*., 2019). In modern building design, HVAC serves as both an installation and control system. However, certain areas within a building, known as under-actuated zones, pose challenges, as the conditions cannot be easily predicted due to undefined occupants or unknown occupancy status (Brooks *et al*., 2015). Unlike controlled zones, where occupant activities are well defined and predictable, under-actuated zones lack clear patterns. This unpredictability emerges from factors such as undefined occupants, variable occupancy status, and the diverse purposes of these spaces. Consequently, the behavior of occupants in these zones is less constrained and more influenced by immediate needs, contributing to fluctuating temperature preferences and occupancy patterns.

Understanding under-actuated zones is pivotal because they introduce distinct challenges for HVAC control strategies. In these zones, traditional predictive methods may fail because of the absence of predetermined occupancy schedules or activities. Occupant actions in these areas can generate varying heat loads and their preferences may change abruptly. Consequently, devising accurate and efficient HVAC control strategies for under-actuated zones requires a comprehensive understanding of the complex interplay between occupant behavior and environmental dynamics. It is crucial to consider the presence and behaviour of occupants in these zones, as their activities can generate heat and impact the cooling load (Wang *et al*., 2018). To address these challenges, Li and Yao (2020) emphasized the significance of assessing occupant presence and behavior in HVAC zones. O'Brien *et al*. (2020) stated that the target building's actual occupant and compressive context-aware data are two important components for optimizing HVAC control. This optimization improves HVAC energy efficiency and enhances occupant comfort (O'Brien *et al*., 2020). Measuring and understanding occupant behavior in under-actuated zones is critical for fine-tuning HVAC control strategies and achieving energy savings. By leveraging this knowledge, building operators can ensure efficient energy utilization and provide occupants with a more pleasant indoor environment.

Previous studies used a stochastic approach to analyze occupant behavior. Three commonly used methods are Bernoulli, Markov chain with agent-based modeling, and survival analysis (Feng *et al*., 2015). Each of these approaches provides distinct advantages and faces specific limitations that shape their applicability in capturing the complexity of occupant behavior. The Bernoulli method focuses on independent and memoryless conditions, where (Holmes and Hassini, 2021) occupants make binary decisions based on fixed probabilities. However, it falls short of capturing the complexity of occupant behavior, influenced by external factors, past experiences, and contextual elements (Yan *et al*., 2015). This simplification may not fully capture the intricate interplay of factors that influence occupant behavior in real-world scenarios, particularly within the context of multifaceted building environments. The Bernoulli process may not be suitable for modeling complex systems with multiple influencing factors (D'Oca *et al*., 2019). Markov chain and agent-based modeling assume that the probability of a situation depends solely on the current state, not past events (Uddin *et al*., 2021; Salimi *et al.*, 2019). Meanwhile, these methods focus on occupant movement patterns and consider previous conditions and activities impacting their decisions (Hong *et al*., 2018; Dziedzic *et al*., 2020), they may not completely encompass the intricacies of behavior influenced by past experiences and external factors. While these models incorporate historical conditions and activities that impact decisions, they may not encompass the entire spectrum of behavior influenced by past experiences and external elements. Thus, while useful for capturing certain aspects of occupant movement patterns, they may not fully capture the complexity of behavior under varying conditions. The Markov Chain method examines the likelihood of occupants transitioning between these states based on observed movement patterns, this helps building operators, and subsequent investigations derive transition probabilities from historical occupancy data. This information aids in predicting and optimizing the use of space, energy consumption, and comfort levels. Agent-based modeling (ABM) simulates individual occupants as autonomous agents, each endowed with unique characteristics, decision-making processes, and preferences. These agents interact with their environment and other individuals, leading to emergent collective behavior at the macroscopic level. For instance, in a workplace setting, individual agents might have specific working hours, preferred meeting locations, or social tendencies that influence their movements in the building. The model captures the probabilities of a person moving from one room to another based on the current state (which room they are in) and the probabilities of different types of movements (e.g., staying in the same room, moving to an adjacent one) (Jia *et al*., 2019; Shelat *et al*., 2020). Overall, these stochastic approaches provide valuable insights into occupant behavior, but their appropriateness depends on the specific complexity and variables of the system being investigated. There is a need to carefully select the appropriate method based on the context and objectives of any research

Survival analysis is a valuable method used to examine the time occupants take to perform specific actions (Denfeld *et al*., 2023; Wang *et al*., 2019). By using this approach, the duration for which occupants perform certain activities can be analyzed (Gunay *et al*., 2016), as well as identify temporal patterns in their behavior, such as daily and weekly cycles, seasonal variations, and overall trends (Barthelmes *et al*., 2018). However, survival analysis methods might not delve into the intricacies of decision-making leading up to an event, a potentially missing context that could provide a more comprehensive understanding of behavior. One significant survival analysis advantage is its ability to reveal inherent temporal patterns in occupant behavior. Multiple research can identify recurrent daily and weekly cycles through modeling and analyzing time-to-event data. For instance, they may observe that the probability of occupants entering a particular room or using specific facilities follows distinct patterns throughout the workweek, with higher and lower utilization during peak and non-peak hours, respectively. These insights

DOI: 10.3844/ajassp.2023.48.64

significantly affect building design, operation, and energy management. Understanding these temporal patterns allows building operators to appropriately align HVAC system operations, lighting schedules, and other services with peak occupancy periods. This, in turn, leads to improved energy efficiency and enhanced occupant comfort. By strategically managing resources and services based on occupant behavioral trends, building owners can optimize facility performance, promote energy conservation, and create a more comfortable and sustainable environment for occupants.

Survival analysis encompassed various methods, including the Kaplan-Meier survivor and Log-rank test (Barthelmes *et al*., 2018). Tree-structured survival models (Wallace, 2014), Survival random forests (Ruyssinck *et al*., 2016; Nurhaliza *et al*., 2022). Deep learning survival models (Dai *et al*., 2020; Yang *et al*., 2019) and Survival regression models (George *et al*., 2014). In contrast to Bernoulli and Markov chain models, which primarily focus on predicting event probabilities at specific time instants (Holmes and Hassini, 2021); survival analysis considers the time taken for an event to occur. Specifically, it examines the duration between the initiation of an event and its completion. In building occupancy and behavior analysis, this methodology is particularly relevant for understanding the time it takes for occupants to perform specific actions, such as adjusting the thermostat, turning lights on or off, or leaving a room (Gunay *et al*., 2016). However, in an under-actuated zone, where occupant's activities are not explicitly controlled, survival analysis becomes valuable in identifying the time it takes for occupants to carry out specific actions in response to environmental conditions. For instance, several researchers can analyze the time taken by occupants to adjust the thermostat concerning changing room temperatures. This information provides valuable insights into occupant comfort preferences and their responsiveness to thermal conditions. Surviving analysis enhances building behavior by going beyond simple event occurrence predictions and exploring its duration aspect, offering a more comprehensive understanding of occupant actions over time. This knowledge empowers building design, operation, and energy management strategies with better insights, ultimately contributing to improved occupant comfort and energy efficiency.

Survival analysis provides diverse methodologies for modeling time-to-event data. These methodologies unveil inherent temporal patterns encompassing diurnal cycles, weekly oscillations, seasonal modulations, and trends within occupant conduct. By dissecting event timing, survival analysis yields insights into how facets such as occupant presence, preferences, and extrinsic conditions influence the chronology of actions. Moreover, survival analysis can be applied in under-actuated zones to study the interaction of occupants with automated systems, such as motion-sensor-activated lighting. Survival analysis with a time-interval paradigm was employed to delve into the nuanced interplay between occupant behavior and HVAC systems within under-actuated zones. By examining the time, it takes occupants to trigger such systems; research is able to assess the effectiveness and user-friendliness of these technologies in enhancing both energy efficiency and occupant comfort. Survival analysis also enables the investigation of time-to-event patterns related to daily and weekly cycles, seasonal variations, and other temporal trends in occupant behavior. Understanding when and how frequently specific events occur allows building designers and operators to optimize system controls and resource allocation. Aligning these strategies with peak occupancy periods leads to more energy-efficient and responsive building environments.

In this research, polynomial regression was employed as one of the Survival regression models to monitor occupant behavior using survival analysis with a time-interval approach. It is a statistical modeling approach that uses data to fit a polynomial function (Sedera and Atapattu, 2019). When assessing their behaviour, the goal was to capture the correlation between time intervals and other factors influencing occupant activities. Polynomial regression was integrated into the survival analysis to create a comprehensive model that considers both cooling load and occupant behavior attributes. The analysis focused on three occupant behavior aspects, namely the number of occupants, their activities, and the electronic devices that are used.

The main objective of this research was to provide valuable insights into cooling load and HVAC systems operation in the under-actuated zone. The justification for using polynomial regression is its ability to capture nonlinear relationships between variables. Fitting a polynomial function to the data made it possible to account for complex interactions and nonlinear trends between time intervals and occupant behavior factors. This approach offered a more accurate representation of the real-world dynamics in the under-actuated zone.

Materials and Methods

This research was conducted at the Universities Trilogy Library, as shown in Fig. 1. The library has a total area of 630.1 m^2 and features a zone that is under and fully actuated. Data were collected from a reading room (room number 5) with eight vents, each covering an area of 25 m^2. The activities of occupants were observed near five vents (C, D, E, F, G), while three of them (A, B, H) located at the crossing into the room were not included in the observations.

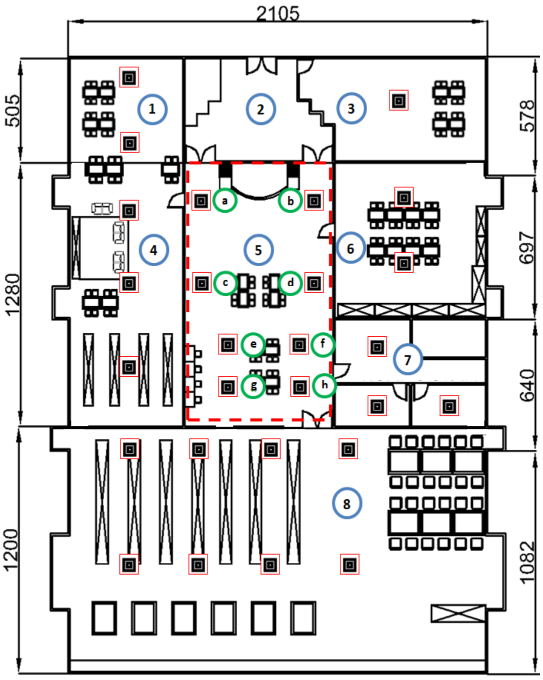


Fig. 1: Layout of Universitas Trilogi Library

The study population and sample included students, lecturers, and academics accessing the library during working hours throughout the active semester, with an average daily attendance of 200. Observations spanned one week, corresponding to five working days and adhering to standard working hours. During this period, occupant behavior was systematically recorded, encompassing the quantification of individuals in each zone and the characterization of activities such as walking, standing, sitting, napping, squatting, and engaging in appliance operations. The data collection process involved a single set of computer equipment, and recording and preliminary processing were performed using Microsoft office excel. Subsequently, an initial data analysis was carried out using R programming studio.

The research investigation consists of four stages. First, occupant behavior data were collected through manual case study observation. Second, the cooling load was calculated based on occupant behavior, activities, and appliance operation. In this scenario, the collected data was used to determine the cooling load attributed to various factors, including occupant behavior, activities, and appliance usage. Analyzing the contributions of each aspect to the cooling load helped to gain a comprehensive understanding of how the actions of the occupants affected it in the reading room. Third, multimodal distribution analysis was conducted using the Hartigan's Dip Test (HDT) and descriptive statistics evaluation of the cooling load data based on different ventilations. Descriptive statistics were computed for the five datasets corresponding to different ventilation. These statistics include mean, median, standard deviation, and range measures.

Analyzing the descriptive statistics of each dataset provided insights into the variability and distribution of cooling load data across various ventilation scenarios, which aided in understanding the cooling dynamics of the room. Fourth, this stage focused on analyzing the appropriate cooling load data time interval using polynomial regression. The present research determined the time interval that accurately fitted the cooling load behavior by calculating the R-squared values and standard error associated with each polynomial regression model. A time interval with a high R-squared value and low standard error indicated a strong relationship between time and cooling load, making it suitable for further analysis. This stage played a crucial role in identifying the most relevant time frame for evaluating the temporal patterns of the cooling load, enabling informed decisions on HVAC system optimization and energy management.

Data Collection

In the present research, the data collection process involved observing occupant behavior in the reading room every 5 min, from 7:45 to 4:45 PM, throughout a week (Monday to Friday) in the active semester. For each vent in the room, three variables were recorded, namely the number of occupants, their activities, and the electronic devices used. The data were saved daily in Microsoft office excel. Each vent had 108 records per day, resulting in a total of 2700 for all five in the reading room over the course of a week. This extensive dataset provides valuable information on occupant behavior and their interactions with electronic devices, enabling a comprehensive analysis of the cooling load and its variability based on different ventilation scenarios.

The decision to conduct observations at 5 min intervals was carefully considered to strike a balance between capturing fine-grained details of occupant behavior and avoiding undue intrusion into their activities. This interval was chosen based on its ability to provide frequent snapshots of the interactions between occupants, electronic devices, and the environment, which are crucial for understanding dynamic cooling load patterns. Furthermore, the observation timeframe from 7:45-4:45 PM was selected to encompass the library's operational hours when the majority of the occupants were present. This time range was reasoned to offer comprehensive insight into occupant behavior throughout the active semester.

Cooling Load Measurement Based on Occupant Behavior

The cooling load attributed to occupant behavior was measured with respect to three significant activities. This includes the number of occupants, their activities, and electronic device usage. Each of these activities contributed to the sensible heat gain, which is a significant factor influencing the cooling load in the reading room. Table 1 shows the sensible heat gain values (expressed in watts per person) for various occupant activities. These values represent the additional heat introduced into the indoor environment due to different types of occupant behavior (Ogedengbe and Erinle, 2015).

Table 1: Zone thermal preferences of occupant activity

No.	Type of activity	Heat transfer coefficient (btu/h)
1	Walk (3.5 Kmph)	307
2	Stand and quiet	307
3	Stand and talk	341
4	Sit and quiet	222
5	Sit and type	256
6	Sit and talk	256
7	Nap	205
8	Squat	222

Table 2: Zone thermal preferences of appliance operation

No.	Type of appliance operation	Heat transfer coefficient (btu/h)
1	Use laptop	854
2	Use tablet	342
3	Use handphone	273
4	Use TV	1025
5	Use computer	1366

For example, sedentary work contributes 60 watts of sensible heat gain per person, while engaging in moderate physical activity adds 120 watts per person. By considering the thermal preferences associated with different occupant activities, the present research aims to comprehensively understand how occupant behavior impacts the cooling load in the reading room. Utilizing these specific heat gain values aids in the analysis of cooling load variations based on occupant activities and devises strategies for optimizing HVAC system operation to maintain a comfortable and energy-efficient indoor environment.

The heat transfer coefficient for each electronic gadget used by the inhabitants throughout their activities was considered in the present research. Table 2 shows the available options for these coefficients, which are determined based on possible appliance operations selected by the occupants (Ogedengbe and Erinle, 2015). These coefficients are directly related to the power consumption of each electrical equipment and serve as a reference for calculating the heat transfer associated with the devices. The sensible heat gains values, along with the corresponding heat transfer coefficients shown in Tables 1-2, respectively, are used to calculate the cooling load contributions resulting from each occupant activity. This comprehensive analysis provides valuable insights into the overall thermal comfort and energy management in the under-actuated zone. By understanding how the activities of the occupants and electronic devices collectively impact the cooling load, the research aims to optimize HVAC system operation and enhance both occupant comfort and energy efficiency.

The dataset was updated to include a new cooling load variable. Subsequently, the cooling load based on occupants was calculated for all records in the dataset. This new cooling load variable was derived by considering three main components. These include cooling load based on occupants (*OCL*), activities (*ActCL*), and appliance operation (*ApCL*). The *OCL* component was determined by multiplying the number of occupants present in the reading room by a cooling load unit of 85 watts or 290 Btu/h per occupant, as per the reference (Kang and Noh, 2019). This component accounts for the thermal energy generated by the occupants and considers their specific cooling requirements. This is the equation of cooling load based on occupant behavior:

$$Total\ CLOB = OCL + ActCL + ApCL \tag{1}$$

$$OCL = (nOcc\ \times OccCoef) + ActCL + ApCL \tag{2}$$

$$ActCL = \sum_{i=0}^{n} nAct_i\ \times\ ActCoef_i \tag{3}$$

$$ApCL = \sum_{i=0}^{n} nAp_i\ \times\ ApCoef_i \tag{4}$$

where:

a) *CLOB* represents the total cooling load based on occupant behavior
b) *OCL* is the cooling load based on the number of occupants present in the reading room
c) *NOcc* is the number of occupants
d) *OccCoef* is the cooling load unit per occupant
e) *ActCL* is the cooling load based on occupant activities
f) $NAct_i$ is the number of occupants engaged in activity *i*
g) $ActCoef_i$ is the sensible heat gain value associated with activity *i*
h) *ApCL* is the cooling load based on electronic device usage
i) NAp_i is the number of instances of electronic device *i* being used
j) $ApCoef_i$ is the heat transfer coefficient associated with electronic device *i*

The *ActCL* was calculated by summing up the total sensible heat gain for each occupant activity present in the reading room. In order to obtain the *ActCL*, the sensible heat gain values from Table 1 (such as sedentary desk work, moderate and light physical activities, standing or walking, use of electronic devices, resting and sleeping) were multiplied by the respective coefficients representing the heat transfer associated with each activity. The result assessed the additional cooling load generated by different occupant activities. Considering the *ActCL* component provides valuable insights into how various occupant activities contribute to the cooling load in the reading room. This comprehensive analysis is crucial for optimizing HVAC system operation and achieving occupant comfort and energy efficiency. The *ApCL* component was determined by considering the cooling load generated by electronic devices and appliances in the reading room. This aspect accounts for the thermal energy emitted by these devices and their impact on the overall cooling load. By including the *APCL* in the analysis, the research gains valuable insights into the impact of electronic devices and appliances on the cooling load in the under-actuated zone. This information is essential for optimizing energy management and ensuring occupant comfort in the indoor environment.

Multimodal Distribution Analysis

This section conducted a descriptive statistical analysis to examine the multimodal distribution of the cooling load data based on occupant behavior in under-actuated zones. The dataset used for this analysis comprised the cooling load values recorded at regular time intervals within the target zones. This analysis aims to gain insights into the distinct patterns of the cooling load and identify potential bimodal characteristics. This sheds light on the underlying factors influencing the cooling load in these areas. Hartigan's Dip Test (HDT), a statistical method designed explicitly for identifying multimodal distributions, was used to assess the presence of bimodal characteristics in the cooling load data. The HDS is used to test the null hypothesis that the data follow an unimodal distribution against the alternative hypothesis of a multimodal one. This test was used to determine whether the cooling load data shows distinct modes or peaks, which could indicate different cooling load regimes or behavior patterns in the under-actuated zones. This analysis provides valuable insights into the cooling dynamics and patterns in these areas, aiding in developing effective HVAC control strategies for optimal energy management and occupant comfort. The following steps outline the procedure for applying the HDS and interpreting its results:

1) Step 1: Calculation of dip statistic values:

    a) $D(F_X)$ calculation: Compute the dip statistic $D(F_X)$, which measures the difference between the empirical Cumulative Distribution Function (eCDF) and the estimated uniform distribution obtained from the cooling load samples. This quantifies the extent of the deviation from unimodality
    b) $D(F_{U_r})$ calculation: Calculate $D(F_{U_r})$, the dip statistic from arbitrary identical samples (bootstrap samples) generated from a uniform distribution. Generate the same number of bootstrap samples as the data points in the original cooling load dataset

2) Step 2: Iterative comparison:

    a) Iterative process: Compare the dip statistic $D(F_X)$ from the original cooling load dataset with each $D(F_{U_r})$ value obtained from the bootstrap samples. This iterative comparison is essential for evaluating the multimodality of the cooling load data
    b) Bootstrap sample comparison: For each bootstrap sample, determine whether $D(F_X)$ is less than $D(F_{U_r})$. If $D(F_X)$ is smaller, it suggests that the cooling load data is closer to an unimodal distribution than a uniform distribution. In such cases, set the indicator function $I_{\Omega_{U_r}}$ for that bootstrap sample to one, indicating unimodality
    c) Multimodal indication: Conversely, if $D(F_X)$ is greater than or equal to $D(F_{U_r})$, it implies that the cooling load data exhibits characteristics of a multimodal distribution. Set $I_{\Omega_{U_r}}$ for that specific bootstrap sample to one, indicating multimodality

3) Step 3: p-value calculation:

    a) Probability assessment: The observed dip statistic $D(F_X)$ is compared with the distribution of $D(F_{U_r})$ values obtained from the bootstrap samples. This comparison calculates the p-value, representing the probability of observing a dip statistic as extreme as $D(F_X)$ under the assumption that the cooling load data follows an unimodal distribution:

$$p-value=\frac{1}{R}\sum_{r=1}^{R}I_{\Omega_{U_r}}(x_i) \tag{5}$$

$$I\Omega\mu_{r=}\begin{cases}1, & D(fx)<D(F\mu_r)\\ 0, & otherwisw\end{cases} \tag{6}$$

    b) Interpretation: A low p-value indicates strong evidence against the null hypothesis of unimodality, suggesting the presence of a significant multimodal pattern in the cooling load data. This statistical approach allows the effective identification and characterization of the cooling load multimodal distribution, providing valuable insights into the underlying cooling load behavior in the under-actuated zones

The Hartigan's dip test results are interpreted based on the calculated p-value. A significant result ($p<0.01$) provides strong evidence of bimodal characteristics in the cooling load data, indicating the presence of distinct cooling load regimes or behavior patterns in under-actuated zones. This significant result provides strong evidence of bimodal characteristics in the cooling load data. To further analyze the multimodal distribution in the cooling load data based on occupant behavior in under-actuated zones, descriptive statistics, such as the mean, mode, standard deviation, skewness, and kurtosis, were calculated. These measures provide a comprehensive distribution analysis, helping to understand the under-actuated zones' cooling load patterns and characteristics. This information is vital for optimizing HVAC system controls and ensuring occupant comfort in such zones.

Time-Interval Analysis

The research investigated various time intervals ranging from 5 min to 2 h (Wang *et al*., 2018). There are five distinct time intervals, namely 5, 15, 30 min, and 1 h, each associated with a different cooling load value. The selected time intervals ranged from very short (5 min) to moderately longer (1 h). This choice allows the analysis to capture the cooling load behavior across various temporal scales. Shorter intervals help to capture rapid changes, whereas longer intervals encompass more gradual variations. The inclusion of 5 min intervals provides a fine-grained view of fluctuations in the cooling load over short periods. This can be important for understanding rapid changes and responses in the cooling system. Time intervals of 15, 30 min, and 1 h are commonly used in practical scenarios, particularly in building and energy management. These intervals align with the common operational decisions and control strategies that building operators and energy managers may implement. Frequently, data collection systems are configured to record values at standard intervals, which are often aligned with the selected time intervals. This ensures the availability of data for analysis without requiring additional adjustments or interpolations.

The intervals were chosen to be diverse, but still statistically significant. Very short intervals might capture more noise than meaningful patterns, whereas excessively long intervals might miss important fluctuations. The selected intervals strike a balance between the granularity and overall trends. Collecting data at extremely short intervals (e.g., seconds) might require higher-frequency data-recording mechanisms, which could be resource-intensive and lead to excessive data volumes. Conversely, intervals much longer than 1 h might not capture important variations in the cooling load. The chosen intervals align with the timeframes over which building energy-management decisions are often made. Polynomial regression was selected as the appropriate method for analyzing nonlinear data within these intervals. The formula of Polynomial regression is stated as follows:

$$f(x) = C_0 + C_1x + C_2x^2 + \cdots + C_nx^n \tag{7}$$

In this equation, $f(x)$ represents the predicted value of the cooling load based on a given time interval x. The equation is a polynomial function, where x is the independent variable (time interval) and $f(x)$ is the dependent variable (cooling load). The coefficients C_0, C_1, C_2,…, and C_n are constants determined by polynomial regression analysis and they play a crucial role in shaping the relationship between the time interval and cooling load. Each coefficient is associated with a particular power x. C_0 represents the intercept, which is the value of the dependent variable when the independent variable (x) is zero. It's the base value of the cooling load. C_1x represents the linear term, where c1 signifies the change in the cooling load for a unit change in the time interval (x). C_2x^2 represents the quadratic term, capturing the curvature of the relationship between the time interval and the cooling load. C_2 influences how the curve bends.

This analysis aimed to determine the most suitable time interval from five available options. In order to achieve this, the R-squared metric was used as a measure of goodness of fit. R-squared is a statistical parameter that assesses the degree to which the independent variable (time interval) explains the variance in the dependent one (cooling load). A high R-squared value indicates a stronger relationship between the time interval and cooling load data. This signifies that the time interval effectively accounts for a significant portion of the cooling load variance. Generally, a desirable goodness of fit is indicated by an R-squared value of not less than 0.9 (Wang *et al*., 2018), depicting a strong relationship between the time interval and cooling load data. The proper time interval was selected by comparing the R-squared values obtained from the regression analysis for each of the five time intervals. The standard error associated with each R-squared value was also considered. The objective was to select the time interval with the highest R-squared value while minimizing the standard error.

Suppose that we calculate the R-squared values for each time interval in Table 3. In this example, the 30 min time interval had the highest R-squared value (0.95). This means that the cooling load variance is explained well by the polynomial regression model using the 30 min time interval.

Table 3: Example of time interval measurement

Time interval	R-squared
5 min	0,92
15 min	0,87
30 min	0,95
1 h	0,82

Based on high R-squared, the 30 min time interval was selected as the most suitable for modeling the cooling load behavior. This hypothetical example illustrates the process of choosing the optimal time interval using the R-squared metric and considering the relationship between the variables, as modeled by the polynomial regression equation. The R-squared metric is particularly useful for analyzing the appropriateness of time intervals, as it provides a quantitative measure of how well the time interval explains the variability in cooling load behavior (Kim, 2021):

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} \tag{8}$$

The R-squared value was calculated by evaluating the sum of squared regression, accounting for the squared differences between the observed cooling load values and the predicted based on the time interval. This value is compared to the sum of squares, representing the squared differences between the cooling load values and their mean. The resulting R-squared value ranges from zero to one.

Results and Discussion

Data were collected from the reading room comprising five observed ventilation areas, with 440 records obtained for each 5 min interval throughout the day. A total of 2200 records were gathered from this room for a week (Monday to Friday). The collected data were saved in a file using Comma-Separated Value (CSV) format, enabling efficient organization and analysis of the data. In this research, data were collected from a reading room that had five different observed ventilation areas. The data were recorded at 5 min intervals, resulting in 440 records for each ventilation area in a single day. Throughout the week (Monday to Friday), 2200 records were obtained from this room. The collected data provides essential information on cooling load data related to occupant behavior, activities, and appliance operation. Further analysis was facilitated by storing the data in a Comma-Separated Value (CSV) format. The data collection process involved gathering information at regular intervals of every 5 min, commencing at 07:45 AM and concluding at 03:00 PM, spanning five weekdays (Monday to Friday) during the active semester. This procedure was conducted for five different ventilation setups, yielding distinct datasets.

Table 4: Hartigan's dip test for 5 vents

Vent	Hartigan's dip value	p-value	Multimodal
C	0.05682	0.0337	Bimodal
D	0.03489	0.6369	Bimodal
E	0.04036	0.3735	Bimodal
F	0.06156	0.0128	Bimodal
G	0.01276	0.0079	Trimodal

The selection of particular days for data collection within the reading room was underpinned by a meticulous rationale driven by the intention to create a dataset that accurately reflects real-world scenarios. The decision to focus on weekdays, spanning from Monday to Friday, was made with the aim of capturing a representative cross-section of a standard work week. During this period, the reading room encountered diverse levels of occupancy, a wide range of activities, and varying environmental conditions. This choice becomes especially meaningful as the week unfolds. Across these weekdays, there was a progression in the dynamics of the reading room. Factors such as occupant density and equipment utilization may display fluctuations that can be instrumental in revealing valuable insights into the variations in cooling load behavior. This temporal evolution reflects the genuine ebb and flow of activities within the reading room during a typical workweek.

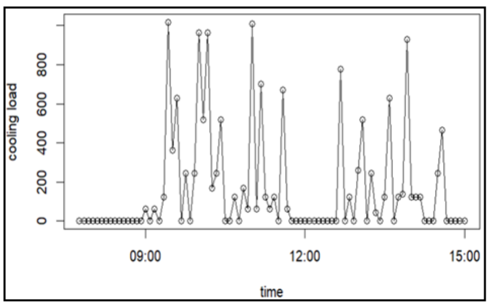
To analyze the underlying distribution characteristics of these datasets, the Hartigan's Dip Test (HDT), a statistical method tailored to identify unimodal or multimodal distributions, was applied. The results of the Hartigan's dip test are shown in Table 4, elucidating the outcomes obtained for each ventilation configuration. Remarkably, the ventilation setups labeled C through F showed compelling evidence to reject the null hypothesis, as reflected in their respective p-values, which were lower than the selected significance level (alpha = 0.05). To comprehensively evaluate the inherent distribution characteristics within our dataset, we used the dip test () function, a prominent component within the R environment sourced from the dip test package. A favored analytical approach, the Hartigan's dip test, was employed to discern potential deviations from the conventional unimodal framework inherent in the datasets. This test entails a comparative assessment that juxtaposes the empirical distribution function with an equivalent median variance endowed with a multimodal distribution function. Moreover, we conducted an exhaustive inquiry to extend the analytical scope to probe the plausible emergence of multimodal patterns intrinsic to our dataset. This exploration encompassed the utilization of the bimodal () and trimodal () functions intrinsic to the dip test package.

Based on the results, it was concluded that these datasets do not follow an unimodal distribution. Interestingly, the dataset corresponding to ventilation G showed a trimodal distribution, indicating the presence of three distinct modes within the data. The identification of a trimodal distribution in the dataset corresponding to ventilation G unveils intriguing insights that warrant further exploration in the context of occupant behavior and cooling load analysis. This distinctive distribution pattern, characterized by the presence of three discernible modes, holds implications that extend beyond the conventional unimodal or even bimodal scenarios. The recognition of a trimodal distribution within the dataset

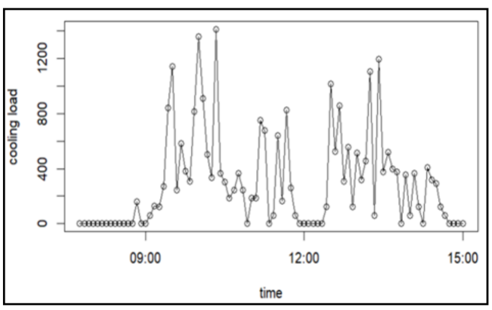
DOI: 10.3844/ajassp.2023.48.64

associated with ventilation G could potentially signify the influence of distinct time-based variations on occupant behavior and cooling load dynamics. Such variations introduce an additional layer of complexity that can significantly impact the interpretation of the observed distribution pattern. The trimodal behavior could be intimately linked to temporal fluctuations in occupant behavior. Each mode may correspond to different time periods, each characterized by unique activities and occupancy levels. For instance, one mode could align with peak occupancy periods, such as during working hours, while another could denote periods of diminished occupancy, potentially during off-peak hours.

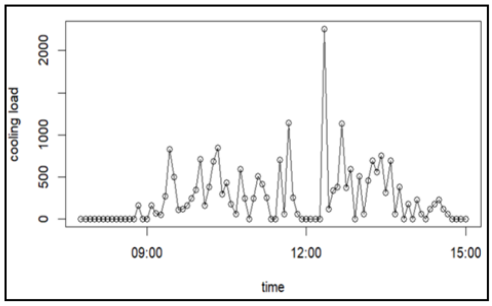
To analyze the cooling load contributions from occupants in the under-actuated zone, three factors were considered, namely occupant number, activities, and appliance loads. The analysis was conducted using a default time interval of 5 min. The findings are shown in Fig. 2 illustrating the cooling load profiles for five different ventilation scenarios. Based on the cooling load measurement shown in Fig. 2. Ventilation F had the highest cooling load, with an average value of 304 Btu/h. This indicates that ventilation F is frequently used for more strenuous activities. It experienced a higher cooling burden on Wednesdays and Fridays, indicating increased occupancy and activity levels. Similarly, ventilation D displayed a significant cooling load with an average value of 295 Btu/h. Ventilation F was subjected to frequent and intense activities. This research reported that ventilation D is most active on Wednesdays and Fridays, as well. Ventilation G had the lowest cooling load, with an average of 138 Btu/h. This implied that this ventilation mode is used less frequently or for less intense activities than F-D.



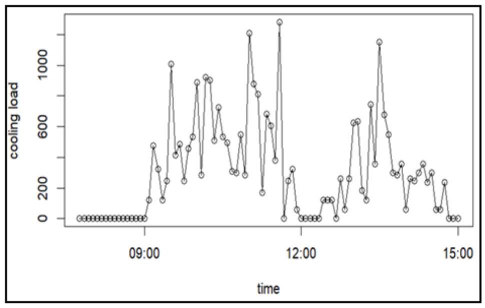
a. Ventilation C



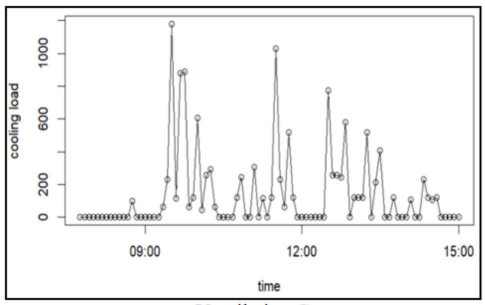
b. Ventilation D



c. Ventilation E



d. Ventilation F



e. Ventilation G

Fig. 2: Cooling load-based occupant for five ventilations

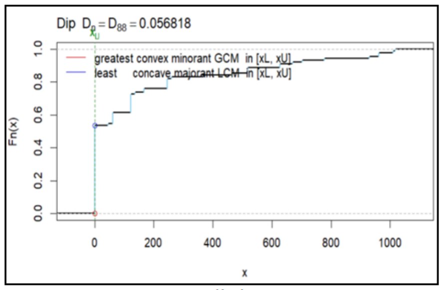
Among the measured modes, Ventilation E had the lowest overall cooling load, but it still showed a relatively high cooling burden, particularly on Fridays. This shows that despite being used less frequently or for less demanding activities, ventilation E experiences a higher cooling load on Fridays. In accordance with the data analysis, two main observations were made. Firstly, F and D were the most frequently used ventilation modes and tended to have higher cooling loads, specifically on Wednesdays and Fridays. Conversely, ventilation G has the lowest cooling burden and is the least used mode. The analysis also revealed that ventilation days exhibit variable usage patterns, with a notably higher cooling load recorded on Fridays. The cooling load attributed to occupants showed a notable pattern, starting at 09:25 AM and exhibiting fluctuations until 12:00 PM. During the lunch break period (12:00-01:00 PM), the cooling load decreased, indicating a reduction in occupant activities. After the lunch break, the cooling load rose again from

DOI: 10.3844/ajassp.2023.48.64

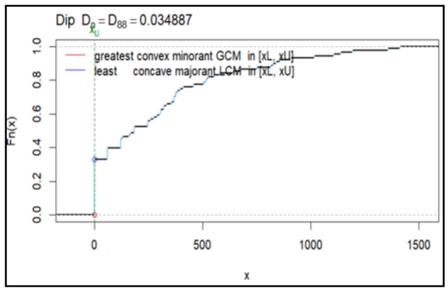
01:00-03:00 PM. This observation suggested that occupants engaged in minimal activities during the lunch break, leading to a decrease in cooling load. When considering all ventilation scenarios, the cooling load was observed to rise above the baseline at 09:00 AM. This indicates that the presence of occupants and their associated activities did not immediately coincide with the opening time of the room. Instead, it took some time for the cooling load to peak as occupants gradually entered and engaged in activities within the reading room throughout the morning. Approximately 1 h after the room's opening, the presence of occupants was felt and a density diagram encompassing all ventilation scenarios was constructed to gain deeper insights. The diagram showed the highest peak at a cooling load value of 0, indicating numerous time intervals with no cooling load contribution. These intervals corresponded to periods when the room remained unoccupied. Specifically, these unoccupied intervals were identified from 07:45-09:00 AM and 12:00-01:00 PM. During these time frames, the cooling load remained at a minimum due to the absence of occupants in the reading room.

Figure 3 shows the cooling load density across all ventilation scenarios over a period. This diagram clearly shows a dominant cooling load value of zero. It indicates numerous time intervals with no cooling load contribution. These intervals likely correspond to periods of low activity or when the reading room is unoccupied. In order to effectively analyze the time variance for ventilations C to F, these were divided into two groups 09:25 to 12:00 PM and 01:00 to 03:00 PM. For ventilation G, three distinct groups were identified 09:25 to 10:30 AM, 10:45 AM to 12:00 PM, and 01:00 to 03:00 PM. This division allows us to examine the cooling load behavior during different time intervals and identify patterns associated with occupant activities and the usage of the reading room.

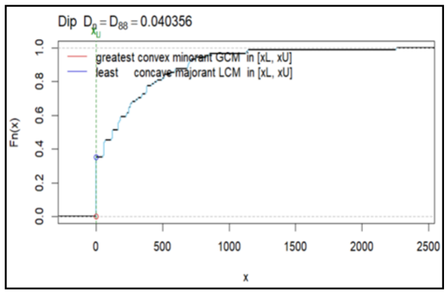
A detailed analysis of temporal cooling load patterns, encompassing room openings, lunch breaks, and peak cooling load times, provides substantial insights into occupant behavior. These patterns align with the existing literature on occupant behavior in similar contexts, confirming the validity of the findings. The findings corroborate the impact of occupant schedules on energy consumption and emphasize the importance of adapting cooling systems to occupant routines. The observed fluctuations during lunch breaks underscore the collective influence of occupant activities on thermal comfort needs, necessitating considerations for transient occupancy-related events in cooling strategies. The identification of peak cooling load intervals aligns with prior research, highlighting the significance of predictive models for effectively managing peak demand. Overall, aligning these patterns with existing literature enriches the validity of the study and contributes to a deeper understanding of the interplay between occupants, cooling loads, and energy management strategies.



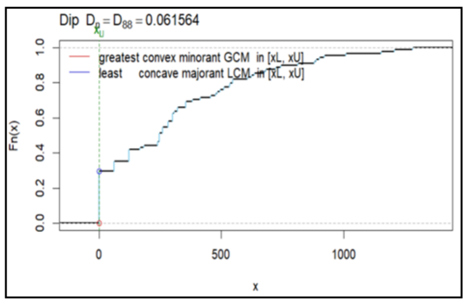
a. Ventilation C



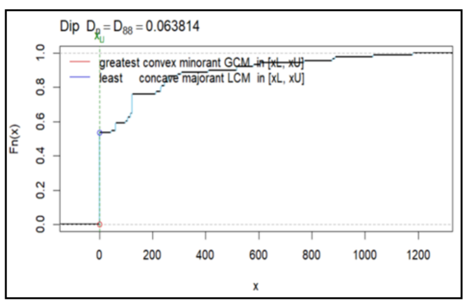
b. Ventilation D



c. Ventilation E



d. Ventilation F



e. Ventilation G

Fig. 3: Density of cooling load based occupant for five ventilations

Table 5: Descriptive statistics for all ventilations

Vent	Mean	Standard error	Median	Standard deviation	Sample variance	Kurtosis	Skewness	Range
C	163,96	29,54	0	275,6	75978	2,66	1,91	1017
D	299,26	37,54	184,2	350,2	122662	1,39	1,38	1412E
	255,35	37,64	122,8	351,1	123282	11,62	2,74	2258
F	308,25	34,47	245,6	321,5	103378	0,69	1,12	1282
G	140,45	26,46	0	246,8	60915	6,01	2,45	1180

Table 6: Descriptive statistics for groups of time variance in ventilations

Vent	Group	Mean	Standard error	Median	Standard deviation	Sample variance	Kurtosis	Skewness	Range
C	1	258,6	57,7	122,8	321,3	103243	0,32	1,23	1009
	2	159,3	49,4	83,6	242,1	58604	3,65	1,96	928
D	1	435,8	69,8	307	389	151420	0,57	1,09	1412
	2	288,1	65	305	318	101410	3,01	1,65	1194
E	1	315,2	51,2	245	285	81395	0,88	1,06	1139
	2	215,1	50,3	122,8	246,5	60805	-0,029	1,09	750
F	1	238,4	60,02	116	334,2	111704	1,85	1,69	1180
	2	91,54	27,96	0	136,9	18765	3,76	1,92	518
G	1	347	111,67	122,8	402,6	162130	-0,21	1,08	1180
	2	183,7	71,72	116	277,79	77171	6,09	2,33	1030
	3	91,5	27,9	0	136,9	18765	3,76	1,92	518

Among the five ventilators used over one week, there were notable differences in cooling load patterns. Table 5 shows a comprehensive overview of the descriptive analysis measurements for all the ventilation systems before being segregated into bimodal and trimodal distributions. These measurements offer valuable insights into the cooling load characteristics of each ventilation. With respect to the five ventilation used during the observation period, each exhibited distinct cooling load patterns. Ventilation C had a moderate mean cooling load of 163.96. However, its relatively high standard deviation of 275.64 indicated a wide spread of cooling load values. The positively skewed distribution (skewness = 1.91) suggested that higher cooling loads were less frequent but could be significantly larger when they occurred. A mode at 0 indicated instances of zero-load or inactive periods for the ventilation system, possibly during low occupancy or reduced cooling loads. The large range of 1017.6 indicated considerable variation in cooling requirements. However, the kurtosis value of 2.66 suggested a peaked distribution with heavier tails, implying the potential presence of outliers or extreme values. This indicated that there were instances with significantly higher cooling loads than the general pattern. Ventilation D had a greater mean cooling load of 299.26 compared to C, indicating a larger average cooling requirement.

Ventilation D also had a larger standard deviation (350.23) and range (1412.8), signifying a broader spread and greater variability in cooling load requirements. The strongly positively skewed distribution (skewness = 1.39) and kurtosis value (1.40) indicated that the cooling load data of Ventilation D deviated from the normal distribution.

Ventilation F had the largest mean cooling load, at 308.29. The standard deviation (321.52) and range (1282.8) demonstrated significant fluctuations in cooling loads, albeit at a lower level than Ventilation E. The moderate right skewness (1.12) indicated a slightly asymmetric cooling load distribution, while the kurtosis value (0.69) showed a nearly normal distribution with relatively lighter tails than those of Ventilation E. Among all the ventilations, G had the lowest mean cooling load of 140.45, indicating the least average cooling requirement. It showed significant variance in cooling loads, although less than Ventilation E, as proven by the standard deviation and range of 246.81 and 1180.6, respectively. The distribution of Ventilation G was strongly positively skewed (skewness = 2.45) with a kurtosis value of 6.01, indicating a peaked distribution with heavier tails compared to a normal distribution. Ventilation E exhibited the highest variability among the five ventilation techniques, with the largest mean cooling load of 255.35 and significant variability reflected by the wide standard deviation and range of 351.12 and 2258, respectively. It also had a highly skewed distribution influenced by high-load occurrences. Ventilation G had the lowest mean cooling load and a skewness similar to E but with a smaller standard deviation and range. The cooling load measurements based on different ventilations revealed variability within the dataset. After a multimodal analysis, the ventilation dataset was divided into several groups. Ventilations C to F were divided into two groups and G has three groups based on the modality analysis. The detailed descriptive statistics of all groups in the ventilations dataset are shown in Table 6. The descriptive statistics for the cooling load at Ventilation C during the entire observation period (07:45-15:00) depict a mean cooling load of 163.96 Btu/h with a standard error of 29.55 Btu/h, reflecting the precision of the sample mean estimate.

The data showed a considerable spread, with a high standard deviation and sample variance of 275.64 Btu/h 75978.02272 Btu^2/h^2. The kurtosis value of 2.66 suggested a distribution with relatively heavy tails, while the positively skewed data (skewness = 1.91) indicated a longer one on the right side. The cooling load range spans 1017.6 Btu/h, illustrating its variability. When comparing the two-time variance groups, significant differences in the cooling load characteristics were observed in groups 1 (09:25-12:00) and 2 (13:00-15:00). Group 1 showed a higher mean cooling load (258.59 Btu/h) compared to the entire observation period. This simply suggested an increased cooling load during that time span. On the other hand, group 2 showed a lower mean cooling load (159.38 Btu/h), indicating it reduced during the later hours. The standard errors for both groups depicted precise mean estimates. Group 1 cooling load variability was higher (standard deviation = 321.32 Btu/h) than the entire observation period, while group 2 showed moderate variability (standard deviation = 242.08 Btu/h). The kurtosis values indicate non-normal distributions, with groups 1 and 2 displaying a more normal-like (kurtosis = 0.32) and heavy-tailed distribution (kurtosis = 3.65), respectively. Skewness values remained positive in both groups, depicting occasional high cooling load values.

These findings have important implications for understanding the temporal patterns of cooling load at Ventilation C. The surge in cooling loads during group 1 (09:25-12:00) was attributed to increased occupancy or more energy-intensive activities during that period. The lower cooling load observed in group 2 (13:00-15:00) was linked to reduced occupancy or less energy-intensive activities in the afternoon. The descriptive statistics for the cooling load at Ventilation D during the entire observation period (07:45-15:00) reveal the following insights, the mean cooling load and standard error are 299.26 Btu/h and 37.55 Btu/h, respectively, indicating the precision of the sample mean estimate. The data showed positive skewness (skewness = 1.39) and moderate kurtosis (kurtosis = 1.40), suggesting a distribution with a longer and relatively moderate tail on the right side compared to the normal distribution. The cooling load variability is evident, with a high standard deviation and range of 350.23 and 1412.8 Btu/h, respectively. By comparing the cooling load between two time variance groups, it was discovered that groups 1 (09:25-12:00) and 2 (13:00-15:00) exhibited distinct patterns. Group 1 showed a significant surge in cooling load, with a higher mean (435.88 Btu/h) and median (307 Btu/h) compared to the entire observation period. The cooling load variability in group 1 was also higher, with a larger and wider standard deviation (389 Btu/h) and range (1412 Btu/h). On the other hand, group 2 showed a decrease in cooling load, which was reflected in lower mean (288.16 Btu/h) and median (305.3 Btu/h) values compared to the entire observation period and group 1. In group 2, the cooling load variability is moderate, with a standard deviation and range of 318.45 and 1194.2 Btu/h, respectively.

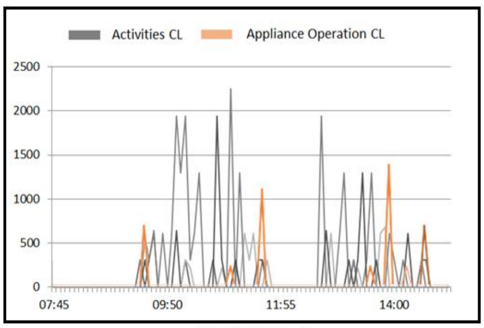
The descriptive statistics of cooling load at Ventilation E for the entire observation period (07:45-15:00) was followed by a comparison between two-time variance groups, namely 1 (09:25-12:00) and 2 (13:00-15:00). During the entire observation period at Ventilation E, the mean cooling load and standard deviation were 255.35 and 37.64 Btu/h, respectively indicating the precision of the sample mean estimate. The mode remained at 0 Btu/h, depicting frequent occurrences of no cooling load. The data showed significant variability, as evident from the high standard deviation and wide range of 351.12 and 2258 Btu/h, respectively. A comparison between the two time variance groups, 1 and 2, revealed distinct patterns. Group 1 experienced a notable surge in cooling load, with a higher mean and median of 315.30 and 245.6 Btu/h, respectively. The cooling load variability in group 1 was relatively moderate compared to the entire observation period. In contrast, Group 2 showed a decrease in cooling load, with a lower mean and median of 215.19 and 122.8 Btu/h. The cooling load variability in group 2 was also moderate. These findings highlighted considerable cooling load variability throughout the day at Ventilation E.

The descriptive statistics of cooling load at Ventilation F during the entire observation period (07:45-15:00) are the mean cooling load and standard error of 308.29 and 34.47 Btu/h, respectively, indicating precise estimation. The data show a considerable spread, with a high standard deviation and sample variance of 321.52 Btu/h and 103378.14 btu^2/h^2. The cooling load distribution exhibited moderately heavy tails (kurtosis = 0.69) and a positively skewed pattern (skewness = 1.12). The median cooling load was 245.6 Btu/h and the mode remained at 0 Btu/h, indicating frequent occurrences of no cooling load during this period. The range of cooling load values spans 1282.8 Btu/h, reflecting substantial variability in cooling load. A comparison between the two-time variance groups, 1 (09:25-12:00) and 2 (13:00-15:00) showed significant differences in cooling load characteristics. During group 1, the mean cooling load was 238.42 Btu/h, lower than the mean of the overall observation period, indicating a reduced cooling load. The standard deviation of 334.22 Btu/h implied relatively higher variability, while the distribution remained positively skewed (skewness = 1.69) with heavier tails (kurtosis = 1.86). The median cooling load in group 1 was 116 Btu/h, further highlighting the decreased cooling load during this period. In group 2, the mean cooling load dropped significantly to 91.54 Btu/h, indicating a substantial decrease in cooling load during the later period. The standard deviation of 136.99 Btu/h reflected moderate variability and the distribution remained positively skewed (skewness = 1.93) with heavy tails (kurtosis = 3.77).

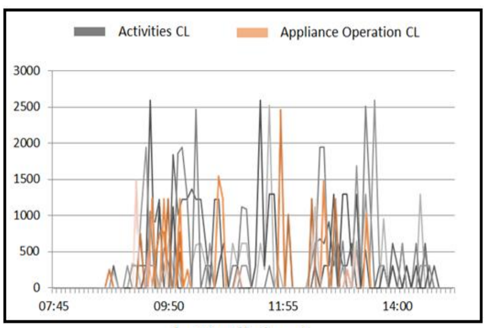
DOI: 10.3844/ajassp.2023.48.64

The cooling load at Ventilation G during the observation period (07:45-15:00) exhibited substantial variability, with a mean and high standard deviation of 140.45 and 246.81 Btu/h, respectively. The data showed frequent occurrences of no cooling load (mode = 0 Btu/h) and a positively skewed distribution with heavy tails (skewness = 2.45, kurtosis = 6.01). When comparing the cooling load among the three-time variance groups, notable differences were observed. Group 1 (09:25-10:30) experienced a significant surge in cooling load, with a mean of 347.03 Btu/h. Group 2 (10:45-12:00) showed an increase in cooling load (mean = 183.79 Btu/h) but to a lesser extent than those in 1. Group 3 (13:00-15:00) depicted a considerably lower cooling load (mean = 91.54 Btu/h) with frequent occurrences of no cooling load.

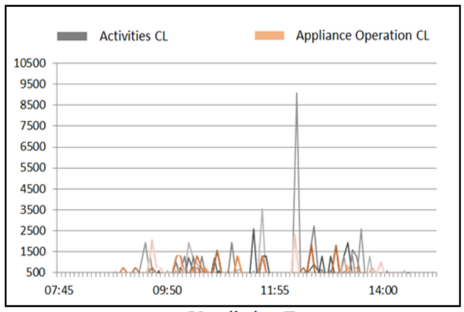
The comparison between the three-time variance groups indicated significant temporal variations in cooling load, with group 1 experiencing an enormous increase in cooling load, group 2 showing an increase to a lesser extent, and group 3 indicating a considerably lower cooling load during the specified period. The distribution of the number of occupants across activities and appliance operation loads is shown in Fig. 4. This provided insights into the connection between occupants and their behavior in the under-actuated zone for all five ventilation days (Monday to Friday). The cooling load-based activities have a higher average value of 176 Btu/h compared to appliance operation, which has an average value of 38 Btu/h. This consistent pattern holds true across all ventilations, indicating that the higher cooling load primarily originated from occupant's activities rather than appliance operations.



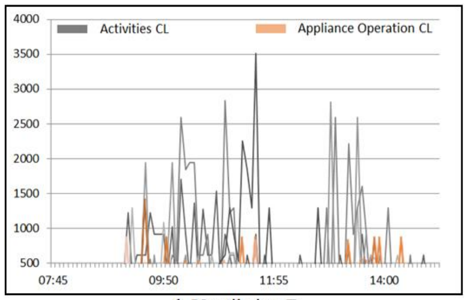
a. Ventilation C



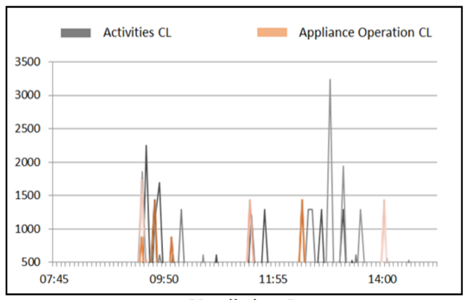
b. Ventilation D



c. Ventilation E



d. Ventilation F



e. Ventilation G

Fig. 4: Cooling load-based activities and appliance operation for five ventilations

A comparison of cooling load patterns across different ventilation techniques constitutes a crucial facet of this study. The comprehensive descriptive statistics provided for each ventilation configuration revealed their distinct cooling load characteristics, providing a foundation for insightful interpretations. However, it is imperative to contextualize these findings within a broader framework of anticipated trends. For instance, in the case of Ventilation C, the moderate mean cooling load of 163.96 Btu/h corresponds to scenarios in which intermittent high-occupancy periods are interspersed with periods of reduced cooling requirements. Ventilation D's higher mean cooling load of 299.26 Btu/h aligns with the notion that increased ventilation rates and higher occupancy densities contribute to elevated cooling demands. Similarly, the highest mean cooling load of 255.35 Btu/h observed in Ventilation E corresponds to its role in accommodating varying occupancies and ventilation needs.

Table 7: Value of R squared for 5 ventilations

Vent	Group	5 min	15 min	30 min	1 h
C	1	0,176	0,232	0,659	1
	2	0,101	0,380	0,993	0,860
D	1	0,239	0,495	0,730	1
	2	0,380	0,753	0,856	0,979
E	1	0,066	0,256	0,562	1
	2	0,492	0,732	0,841	1
F	1	0,197	0,481	0,879	1
	2	0,424	0,698	0,861	1
G	1	0,332	0,804	1	1
	2	0,221	1	1	1
	3	0,177	0,541	0,786	1

The characteristics of Ventilation F, with a mean of 308.29 Btu/h, show a balanced distribution indicative of a ventilation system designed to cater to average occupancy density and thermal comfort requirements. The lowest mean cooling load of 140.45 Btu/h in Ventilation G reflects its function as an energy-efficient ventilation strategy that is suitable for scenarios with reduced cooling needs. This contextualization bridges the empirical findings with theoretical expectations, enhancing the discussion by situating the results in a broader context of ventilation characteristics and occupant behaviors. A significant aspect of this investigation was determining the optimal time interval for data collection and modeling. Striking a balance between capturing detailed information within a manageable time frame and avoiding excessive data volume, which could complicate proper quantification, was crucial. However, a longer period could introduce uncertainties during the validation process. Finding the ideal time interval ensured both accuracy and efficiency during the analysis.

In order to address this issue, the obtained data were fitted and modeled over multiple time intervals 5, 15, 30 min, and 1 h. The aim was to investigate how different temporal resolutions would impact the modeling results. Table 7 shows a detailed summary of the cooling-load-based occupant data distribution across each of the five time intervals, 5, 15, 30 min, and 1 h. It includes information on occupant number, activity, and appliance loads, allowing a comprehensive assessment of data variability for modeling applications. It also presents the coefficients of fit for the cooling-load-based occupant, indicating the level of agreement between the collected data and the predicted result. Analyzing these coefficients helps to ascertain the accuracy and reliability of the models built using each time interval. The careful selection of the time interval used throughout the data fitting process is crucial for producing relevant and robust modeling outcomes. By assessing the coefficients of fit and evaluating the distribution of cooling-load-based occupant data, informed decisions could be made regarding the most appropriate time interval for modeling purposes. This ensures that the models capture the underlying patterns and relationships in the data accurately, leading to meaningful and precise results. Table 7 shows the proper time intervals for different groups within each ventilation scenario. In ventilation C, groups 1-2 had proper time intervals of 1 hand 30 min, respectively. For ventilation D, E, and F, the proper time interval for all groups was 1 h, indicating similar cooling load averages for occupant number, activities, and appliance operation in these areas. However, ventilation G exhibited more variation, with different proper time intervals for each group. Groups 1, 2, and 3 had proper time intervals of 30, 15 min, and 1 h, respectively. These differences suggest distinct cooling load characteristics and occupant behavior for each group within ventilation G.

The proper time intervals in the different ventilation groups tend to vary, with some having shorter intervals (30 and 15 min) and others having longer ones (1 h). The variations in proper time intervals were due to the specific needs of the occupants and activities in each area. Tailoring the time intervals for data collection and modeling based on these unique characteristics is crucial for accurately representing and analyzing cooling load patterns. This approach ensures that the modeling results are relevant and reliable for each ventilation scenario.

The exploration of diverse time intervals for data fitting and modeling, as exemplified in this study, offers valuable insights into the intricate dynamics of cooling load patterns across varying temporal resolutions. The analysis of R-squared values across different intervals, as presented in Table 6, serves as a quantifiable metric to assess the suitability of each interval for modeling. However, delving into the rationale underlying the selection of specific time intervals and expounding on the profound implications of these choices on model accuracy, robustness, and practical applicability, provides a more comprehensive understanding of this methodological approach.

The rationale behind the time interval selection resides in the pursuit of accurately capturing the inherent variability and transient behaviors within the cooling load patterns. Shorter intervals, such as 5 min, facilitate the capture of rapid fluctuations and nuanced occupant activities; however, they might also amplify noise and obscure overarching trends. In contrast, longer intervals, such as 1 h, offer a broader perspective on cooling load trends but might smooth out important transient variations. Intermediate intervals, such as 15 and 30 min, strike a balance between these extremes. The crux of this matter is that the chosen time interval hinges on research goals, occupant behaviors, and the desired temporal resolution in modeling. The consequences of this choice extend to the accuracy and applicability of models. Opting for shorter intervals may result in models that are sensitive to minor variations but are less capable of extrapolating to longer time frames. Conversely, models based on longer intervals might provide better generalization but could miss rapid changes. The effectiveness of the models during transition periods or sudden shifts in occupant behavior hinges on the chosen interval. Moreover, the practical utility of these models in real-world scenarios is at stake. Overly short intervals could render models susceptible to noise, whereas overly long intervals could compromise their responsiveness to adaptive control strategies.

In conclusion, selecting a time interval for data fitting and modeling is a nuanced decision based on research goals and temporal intricacies, dictating the accuracy, reliability, and real-world utility of the models and providing a comprehensive methodological foundation for their implementation and interpretation in practical contexts.

Conclusion

In conclusion, this research focused on three important characteristics of tenant behavior, the number of occupants, their activities, and the use of technological devices. Analyzing real-world occupant data from a library, provided useful insights into the dynamic nature of occupant activity and its impact on cooling load. In order to have a better understanding of the subtle aspects of occupant behavior in under-actuated zones, survival analysis was used to discover temporal patterns, such as daily and weekly cycles, as well as seasonal fluctuations. The incorporation of polynomial regression enabled the capture of nonlinear interactions between time intervals and occupant behavior parameters. This comprehensive approach provided a more accurate picture of the complex dynamics within such zones, with R-squared values greater than 0.8, indicating a robust association. The present research successfully determined the optimal time for monitoring cooling load changes based on occupant behavior, which could be valuable for enhancing HVAC control strategies. To expand the applicability of the findings, future research needs to encompass a broader range of under-actuated zones and building contexts. This broader approach would help to better understand occupant behavior dynamics and their implications for HVAC management systems.

Acknowledgment

We would like to express our sincere gratitude to all those who contributed to the successful completion of this research. First and foremost, we extend our heartfelt appreciation to our supervisor at Asia e University (AeU) and Universiti Kuala Lumpur (UniKL), Assoc. Prof. Aedah Binti Abd Rahman and Ts. Dr. Amna Saad, for their invaluable guidance, encouragement, and insightful feedback throughout the project. We would also like to thank Universitas Trilogi for giving permission to use the library as the object of research.

Funding Information

We are indebted to the Department of Higher Education, Ministry of Education and Culture of Indonesia for providing funding to conduct this research, without which this study would not have been possible.

Author’s Contributions

Yaddarabullah: Contributed to the study design, and conceptualization of the study, coordinated the research activities, played a significant role in manuscript preparation, performed data analysis and critically revised the manuscript for important intellectual content.

Yusuf Maulana Akbar: Conceived and designed the study, collected and analyzed the data, wrote the initial draft of the manuscript, conducted a literature reviewed and contributed to the interpretation of the results.

Aedah Binti Abd Rahman: Participated in data interpretation, critically reviewed the manuscript and provided important insights into the discussion section.

Amna Saad: Provided guidance on statistical analysis, interpreted the data, and critically reviewed the manuscript.

Ethics

This research study was conducted in accordance with the ethical principles outlined by the department of higher education, ministry of education and culture of Indonesia, and in compliance with relevant international regulations. The authors declare that there is no conflict of interest that could have influenced the conduct or reporting of this research. The data from this study can be obtained by contacting the corresponding author directly via email.

References

Bahramnia, P., Hosseini Rostami, S. M., Wang, J., & Kim, G. J. (2019). Modeling and controlling of temperature and humidity in building heating, ventilating, and air conditioning systems using model predictive control. *Energies*, *12*(24), 4805. https://doi.org/10.3390/en12244805

Barthelmes, V. M., Li, R., Andersen, R. K., Bahnfleth, W., Corgnati, S. P., & Rode, C. (2018). Profiling occupant behaviour in Danish dwellings using time use survey data. *Energy and Buildings*, *177*, 329-340. https://doi.org/10.1016/j.enbuild.2018.07.044

Brooks, J., Kumar, S., Goyal, S., Subramany, R., & Barooah, P. (2015). Energy-efficient control of under-actuated HVAC zones in commercial buildings. *Energy and Buildings*, *93*, 160-168. https://doi.org/10.1016/j.enbuild.2015.01.050

Dai, X., Liu, J., & Zhang, X. (2020). A review of studies applying machine learning models to predict occupancy and window-opening behaviors in smart buildings. *Energy and Buildings*, *223*, 110-159. https://doi.org/10.1016/j.enbuild.2020.110159

Denfeld, Q. E., Burger, D., & Lee, C. S. (2023). Survival analysis 101: An easy start guide to analysing time-to-event data. *European Journal of Cardiovascular Nursing*, *22*(3), 332-337. https://doi.org/10.1093/eurjcn/zvad023

D'Oca, S., Gunay, H. B., Gilani, S., & O'Brien, W. (2019). Critical review and illustrative examples of office occupant modelling formalisms. *Building Services Engineering Research and Technology*, *40*(6), 732-757. https://doi.org/10.1177/0143624419827468

Dziedzic, J. W., Yan, D., Sun, H., & Novakovic, V. (2020). Building occupant transient agent-based Model-Movement module. *Applied Energy*, *261*, 114417. https://doi.org/10.1016/j.apenergy.2019.114417

Feng, X., Yan, D., & Hong, T. (2015). Simulation of occupancy in buildings. *Energy and Buildings*, *87*, 348-359. https://doi.org/10.1016/j.enbuild.2014.11.067

George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, *21*, 686-694. https://doi.org/10.1007/s12350-014-9908-2

Gunay, H. B., O'Brien, W., & Beausoleil-Morrison, I. (2016). A toolkit for developing data-driven occupant behaviour and presence models. In *Proceedings of the IBPSA-Canada eSim Conference*. http://www.ibpsa.org/?page_id=662

Holmes, J., & Hassini, S. (2021). Discrete-time Markov chain modelling of the ontario air quality health index. *Water, Air, and Soil Pollution*, *232*, 1-13. https://doi.org/10.1007/s11270-021-05096-1

Hong, T., Chen, Y., Belafi, Z., & D'Oca, S. (2018). Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs. *In Building Simulation* (Vol. *11*, pp. 1-14). Springer Berlin Heidelberg. https://doi.org/10.1007/s12273-017-0396-6

Jia, M., Srinivasan, R. S., Ries, R., Weyer, N., & Bharathy, G. (2019). A systematic development and validation approach to a novel agent-based modeling of occupant behaviors in commercial buildings. *Energy and Buildings*, *199*, 352-367. https://doi.org/10.1016/j.enbuild.2019.07.009

Kang, Y. J., & Noh, Y. (2019). Development of Hartigan's dip statistic with bimodality coefficient to assess multimodality of distributions. *Mathematical Problems in Engineering*, *2019*, 1-17. https://doi.org/10.1155/2019/4819475

Kim, M. S. (2021). Comparative R squared. *Available at SSRN 3790200*. https://doi.org/10.2139/ssrn.3790200

Li, X., & Yao, R. (2020). A machine-learning-based approach to predict residential annual space heating and cooling loads considering occupant behaviour. *Energy*, *212*, 118676. https://doi.org/10.1016/j.energy.2020.118676

Nurhaliza, S., Sadik, K., & Saefuddin, A. (2022). A comparison of cox proportional hazard and random survival forest models in predicting churn of the telecommunication industry customer. *Barekeng*: *Jurnal Ilmu Matematika dan Terapan*, *16*(4), 1433-1440. https://doi.org/10.30598/barekengvol16iss4pp1433-1440

O'Brien, W., Wagner, A., Schweiker, M., Mahdavi, A., Day, J., Kjærgaard, M. B., ... & Berger, C. (2020). Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment*, *178*, 106738. https://doi.org/10.1016/j.buildenv.2020.106738

Ogedengbe, T. I., & Erinle, T. J. (2015). Computer-aided load estimation of air-conditioning systems for non-residential and residential buildings.

Papadopoulos, S., Kontokosta, C. E., Vlachokostas, A., & Azar, E. (2019). Rethinking HVAC temperature setpoints in commercial buildings: The potential for zero-cost energy savings and comfort improvement in different climates. *Building and Environment*, *155*, 350-359. https://doi.org/10.1016/j.buildenv.2019.03.062

Ruyssinck, J., van der Herten, J., Houthooft, R., Ongenae, F., Couckuyt, I., Gadeyne, B., ... & Dhaene, T. (2016). Random survival forests for predicting the bed occupancy in the intensive care unit. *Computational and Mathematical Methods in Medicine*, *2016*. https://doi.org/10.1155/2016/7087053

Salimi, S., Liu, Z., & Hammad, A. (2019). Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain. *Building and Environment*, *152*, 1-16. https://doi.org/10.1016/j.buildenv.2019.01.052.

Sedera, D., & Atapattu, M. (2019). Polynomial regression and response surface methodology: Theoretical non-linearity, tutorial and applications for information systems research. *Australasian Journal of Information Systems*, *23*, 1-35. https://doi.org/10.3127/ajis.v23i0.1966

Shelat, S., Daamen, W., Kaag, B., Duives, D., & Hoogendoorn, S. (2020). A Markov-chain activity-based model for pedestrians in office buildings. *Collective Dynamics*, *5*, 423-430. https://doi.org/10.17815/CD.2020.78

Uddin, M. N., Wei, H. H., Chi, H. L., & Ni, M. (2021). Influence of occupant behavior for building energy conservation: A systematic review study of diverse modeling and simulation approach. *Buildings*, *11*(2), 41. https://doi.org/10.3390/buildings11020041

Wallace, M. L. (2014). Time-dependent tree-structured survival analysis with unbiased variable selection through permutation tests. *Statistics in Medicine*, *33*(27), 4790-4804. https://doi.org/10.1002/sim.6261

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, *51*(6), 1-36. https://doi.org/10.1145/3214306

Wang, Z., Ding, Y., Deng, H., Yang, F., & Zhu, N. (2018). An occupant-oriented calculation method of building interior cooling load design. *Sustainability*, *10*(6), 1821. https://doi.org/10.3390/su10061821

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

Yang, S., Ma, W., Pi, X., & Qian, S. (2019). A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources. *Transportation Research Part C*: *Emerging Technologies*, *107*, 248-265. https://doi.org/10.1016/j.trc.2019.08.010