IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 July 2024, accepted 26 July 2024, date of publication 1 August 2024, date of current version 9 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436940

RESEARCH ARTICLE

A Study of Temporal Correlation Between Space Utilization and Electricity Consumption in Buildings Using Wi-Fi Probe Data

SUNSIKA CHAIKUL[1,2], YOTTANA KHUNATORN[3,4], (Member, IEEE), AND SANTI PHITHAKKITNUKOON[1,2]

[1]City Context Laboratory, Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Excellence Center in Infrastructure Technology and Transportation Engineering (ExCITE), Chiang Mai University, Chiang Mai 50200, Thailand
[3]Department of Mechanical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand
[4]Energy Research and Development Institute-Nakornping, Chiang Mai University, Chiang Mai 50200, Thailand

Corresponding author: Santi Phithakkitnukoon (santi@eng.cmu.ac.th)

This work was supported in part by Chiang Mai University.

ABSTRACT This paper presents an initial exploration into understanding the intricate relationship between space utilization and electricity consumption within buildings. Leveraging Wi-Fi probe data for physical space utilization in 10 buildings at the Faculty of Engineering, Chiang Mai University, Thailand, the study introduces a novel approach to comprehend space utilization. Through the analysis of factors inferred from the Wi-Fi data including crowdedness, mobility, and entropy, correlated with changes in electricity consumption, the research aims to reveal strategies for optimizing real-time energy use. The temporal relationship is scrutinized by examining correlations between each space utilization attribute and electricity consumption at various time lags. Fluctuations in space utilization levels offer valuable insights into anticipated energy consumption in the near future. Results, in general, indicate that the 30–45 minutes timeframe serves as an early indicator for anticipated increases in electricity usage when considering crowdedness, while the 15–30 minutes timeframe is indicative when mobility and entropy are concerned. Buildings were clustered together based on electricity usage profiles and space allocation, yielding intuitive results. By delving into the dynamic interplay between space utilization and electricity consumption, the study contributes valuable insights to discussions on sustainable building practices and energy management.

INDEX TERMS Urban informatics, energy, building, temporal correlation, space utilization, electricity consumption, Wi-Fi data analysis.

I. INTRODUCTION

The efficient utilization of energy resources has become a paramount concern in the face of increasing energy demand and the pressing need to address environmental sustainability. In this context, buildings stand as significant contributors to energy consumption, particularly in urban areas where they account for a substantial portion of the total energy usage [1]. Within the realm of building energy consumption, the interplay between space utilization patterns and electricity consumption has emerged as a critical area of investigation [2], [3], [4]. Understanding the temporal correlation between how spaces are used within buildings and the corresponding electricity consumption holds the potential to unveil opportunities for energy optimization, design enhancements, and policy interventions.

In recent years, advancements in sensing technologies, data analytics, and building management systems have facilitated the collection and analysis of comprehensive occupancy and energy usage data within buildings [5], [6]. This wealth of data opens the door to unraveling the intricate dynamics that link space utilization patterns and electricity

The associate editor coordinating the review of this manuscript and approving it for publication was Renato Ferrero.

consumption [7]. However, despite the growing body of research in this domain, several questions remain unanswered, demanding a systematic exploration of the underlying causal mechanisms.

The link between space utilization and energy consumption in buildings has garnered considerable attention from researchers, architects, urban planners, and policymakers alike. Studies have underscored the potential for significant energy savings by aligning building operations with actual space occupancy patterns. For instance, Erickson et al. [8] demonstrated that occupancy modeling can lead to more effective scheduling of heating, cooling, and lighting systems, resulting in substantial reductions in energy consumption. Furthermore, Latha et al. [9] revealed a direct correlation between spatial configuration and energy demand, emphasizing the importance of layout design in determining lighting and cooling requirements. This suggests that the physical arrangement of spaces within a building can exert a causal influence on energy usage, an aspect that requires further exploration in diverse building types and contexts. In parallel, the emergence of smart building technologies has enabled the real-time monitoring of space occupancy and energy consumption. Through advanced sensor networks and data analytics, Tien et al. [10] demonstrated how occupancy-driven lighting and climate control adjustments can lead to energy efficiency improvements. These findings highlight the potential of dynamic space utilization management in influencing energy demand patterns. Moreover, a study by Gui [11] examined how various types of space usage in university buildings impact energy consumption. They found that wet laboratories and health-related areas use more energy due to equipment and ventilation needs. While research-focused buildings have higher energy use, increasing teaching space has a bigger energy impact.

Despite these insights, the establishment of a temporal correlation between space utilization and electricity consumption remains uninvestigated. Traditionally, assessing space utilization has been reliant on manual observations, surveys, or sensor-based systems, each with its own limitations and complexities. The proliferation of Wi-Fi networks within built environments offers an innovative avenue for capturing space utilization patterns in a non-intrusive and cost-effective manner. Motivated by our previous study in utilizing Wi-Fi probe data for physical space segmentation [12], this work delves into the innovative use of Wi-Fi connectivity data as a tool for sensing and understanding space utilization within buildings.

Beyond its role in occupancy estimation, this study extends its scope to investigate the temporal correlation between space utilization and electricity consumption in buildings. Recognizing that the manner in which spaces are used directly affects energy demand for lighting and cooling, exploring this link provides essential insights for sustainable building operations. By examining whether variations in space utilization coincide with changes in electricity consumption, the research seeks to uncover potential strategies for optimizing energy use based on real-time space dynamics.

Through this dual exploration of using Wi-Fi connectivity data for space utilization analysis and understanding its interplay with electricity consumption, this study addresses vital concerns at the intersection of building management and energy efficiency. The findings hold the potential to inform smarter and more adaptive building strategies, ushering in a new era of energy-conscious design and operational practices.

The subsequent sections of this paper will detail the methodology employed, the dataset utilized, and the analytical framework constructed to discern the temporal correlation between space utilization and electricity consumption. Through an integrated approach that merges quantitative analysis with real-world insights, we aspire to contribute to the evolving discourse on sustainable building practices and energy management.

II. METHODOLOGY

A. DATASET

This study was carried out based on two main data sources collected from 10 academic buildings (total area of 28,224 m^2) within the Faculty of Engineering, Chiang Mai University, Thailand. Both data were collected over the period from 9th January – 3rd February 2020. Electricity consumption data was collected from the energy usage meters, which includes a building ID, energy consumption in kWh and its corresponding timestamp with 15-minute sampling rate.

For space utilization, we take an opportunistic sensing approach making use of data that has already been collected for one purpose, but it can be used to opportunistically for another purpose. In this case, a Wi-Fi network connectivity data recorded for network performance monitoring and planning was used as a proxy for physical space utilization. The data was collected from 97 Wi-Fi access points (APs) belonging to these 10 buildings with a 15-minute sampling rate. Each record includes a connected device ID, AP ID, AP's geolocation (latitude and longitude), building ID, and its corresponding timestamp.

For our analysis, the data of each i^{th} AP can be denoted as D_i, as follows.

$$D_i = \left\{id, lat, lon, \{d_j\left(timestamp, \left\{Dev_j^i\right\}\right) \mid j = 1, 2, 3, \ldots, z_i\}\right\}, \quad (1)$$

where id represents the unique identifier of the AP, lat and lon denote the latitude and longitude coordinates respectively of the AP, d_j signifies the connectivity log at the j^{th} timestamp, Dev_j^i represents the set of device IDs connected to the i^{th} AP at the j^{th} timestamp, while z_i indicates the total count of unique timestamps.

The locations of the buildings and APs considered in this study are shown in Fig. 1, where the corresponding number of APs in each building and its space area information are given in Table 1. These buildings are multi-purpose buildings with rooms allocated for different activities e.g., lecturing,

TABLE 1. Building's information including the number of access points, space area, and the percentage of room types in the building.

Building	No. of APs	Area (m^2)	Room types Lecture	Meeting	Research lab	Student club	Academic office	Admin office	Canteen
B1	2	1,653	70	0	0	0	0	30	0
B2	6	5,768	40	0	20	10	15	15	0
B3	6	1,424	100	0	0	0	0	0	0
B4	3	940	0	0	0	0	50	50	0
B5	48	8,956	45	0	25	0	15	10	5
B6	4	1,219	0	0	0	30	0	0	70
B7	8	2,516	60	0	0	0	0	40	0
B8	9	2,568	70	0	0	30	0	0	0
B9	9	1,826	50	50	0	0	0	0	0
B10	2	1,354	60	40	0	0	0	0	0

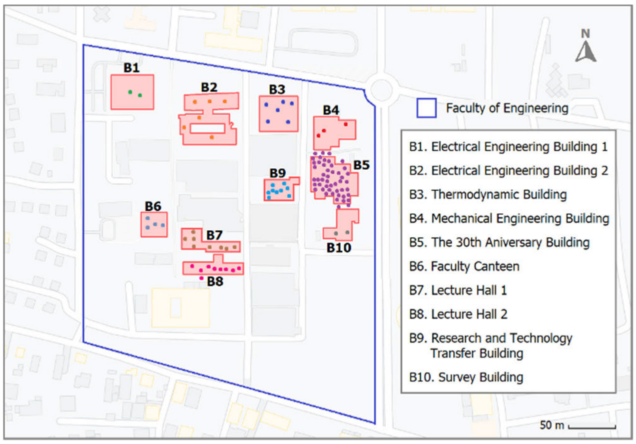
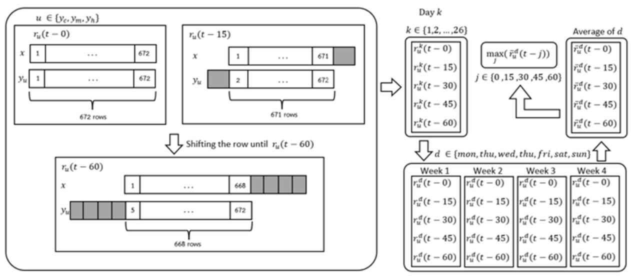


FIGURE 1. Locations of the buildings and Wi-Fi access points considered in this study. Access point colors are used to differentiate their belonging to different buildings. Geocoordinates of the upper left and bottom right corners are 18.797390, 98.949412 and 18.793167, 98.955762, respectively.

meeting, research, and so on. As these different room types may reflect on how the building is utilized, Table 1 also includes the estimated percentage of different room types in each building.

B. ANALYSIS FRAMEWORK

In conducting a temporal correlation analysis to explore the relationship between space utilization and electricity consumption in buildings, Time-lagged Cross-correlation (TLCC) [13] was chosen as a rational approach in our study. TLCC is a statistical method used to identify and measure the degree of similarity between two time series datasets as one is shifted in time relative to the other. This technique is particularly valuable in scenarios where the goal is to uncover potential lead-lag relationships, which can indicate causal effects or predictive patterns. TLCC has been extensively applied in various fields to analyze temporal dependencies [14], [15], [16]. By shifting one time series relative to another and calculating the correlation at each step, TLCC helps identify the time delay at which the correlation between the two series is maximized. This is critical for understanding whether changes in one variable (e.g., electricity consumption) precede changes in another variable (e.g., space utilization), thereby suggesting a possible causal relationship. The overview of our temporal correlation examination between the electricity consumption and the space utilization in each building is shown in Fig. 2.



Notation	Description
x	Vector representing the building's electricity usage profile (vector of size 4x24x7 = 672)
y_c	Vector representing the building's crowdedness (vector of size 4x24x7 = 672)
y_m	Vector representing the building's mobility (vector of size 4x24x7 = 672)
y_h	Vector representing the building's entropy (vector of size 4x24x7 = 672)
$r_u(t-j)$	R-squared value, representing the correlation between the electricity usage and one of the space utilization attributes at j minutes prior, where $u \in \{y_c, y_m, y_h\}$.
$\overline{r_u^d}$	Average r-squared value to portray the observed correlation for each day of the week, where d = Monday, Tuesday, Wednesday, …, Sunday
$\overline{r_u^k}$	Average r-squared value to portray the observed correlation for each day in the observation period, where k = 1, 2, 3, …, 26

FIGURE 2. Overview of the temporal correlation examination between the electricity consumption and the space utilization attributes in each building.

As described in the literature, the utilization of physical space is often defined by two key factors: the density of individuals engaged in activities within the space, resulting in varying degrees of *crowdedness* [17], [18], and the movement of people, which generates dynamism and different levels of *mobility* within that space [19], [20]. Utilizing Wi-Fi data, we determined the degree of crowdedness for each AP as the maximum number of unique device IDs connected simultaneously within a defined time period, denoted as T. In this study, T was set to 15 minutes so that it aligns with the electricity consumption's sampling rate. If C_i represents a set of crowdedness values for the i^{th} AP across multiple time periods, it can be defined as follows.

$$C_i = \{c_1, c_2, c_3, \ldots, c_t, \ldots, c_N\} \tag{2}$$

where c_t represents the crowdedness observed during the t^{th} time period, while N denotes the total number of periods. In our analysis, employing a 15-minute interval (T), the average crowdedness value was calculated for each quarter-hour segment within an hour, across each hour of the day, and for every day of the week. Denoted as C'_i, this set encompasses these average values, defining the crowdedness attributes specific to the AP, as follows.

$$C'_i = \{c'(q, h, d) | q \in Quarters, h \in Hours, d \in Days\}, \tag{3}$$

where $c'(q, h, d)$ is the average crowdedness observed during the quarter q of hour h of day d, where $Quarters = \{1, 2, 3, 4\}$, $Hours = \{0, 1, 2, 3, \ldots, 23\}$, and $Days = \{$Monday, Tuesday, Wednesday, …, Sunday$\}$. Consequently, the set C'_i comprises $4 \times 7 \times 24 = 672$ members, ranging from $c'(1, 0, Monday)$ to $c'(4, 23, Sunday)$.

Similarly, the mobility level for each AP is defined by tallying the total number of device ID connections and disconnections within a time period T. These instances of AP connection and disconnection are regarded as indicators of individuals entering and leaving the AP, intuitively reflecting mobility patterns. Representing the set of mobility values for the i^{th} AP across multiple time periods, M_i can be defined as follows.

$$M_i = \{m_1, m_2, m_3, \ldots, m_t, \ldots, m_N\}, \tag{4}$$

where m_t represents the mobility observed during the t^{th} period, while N is the total number of periods. Similar to the crowdedness value, using a 15-minute interval (T), the mobility attribute is calculated as a collection of average mobility values across each quarter-hour segment within an hour, for every hour of the day, and for each day of the week, as follows.

$$M'_i = \{m'(q, h, d) | q \in Quarters, h \in Hours, d \in Days\}, \tag{5}$$

where $m'(q, h, d)$ is the average mobility during the quarter q (where $q = 1, 2, 3, 4$) of hour h (where $h = 0, 1, 2, 3, \ldots, 23$) of day d (where $d =$ Monday, Tuesday, Wednesday, …, Sunday).

While individual human behavior might seem unpredictable, recognizable patterns and routines often emerge. These collective behavioral patterns become more discernible when viewed over time, revealing varying degrees of randomness. Shannon's entropy, a measure of uncertainty or randomness in human behavior [21], proves relevant in analyzing Wi-Fi connectivity behavior in our context. Hence, apart from crowdedness and mobility, connectivity entropy was derived from the data as another feature defining physical space characteristics. The entropy computations aimed to gauge the level of randomness in connectivity observed during each quarter-hour segment within an hour, across every hour of the day, and for each day of the week. Thus, a set of connectivity entropy values (H'_i) for the i^{th} AP can be defined as follows.

$$H'_i = \{H_i(X(q, h, d)) | q \in Quarters, h \in Hours, d \in Days\}, \tag{6}$$

where $H_i(X(q, h))$ is calculated according to the Shannon's entropy [21] as follows.

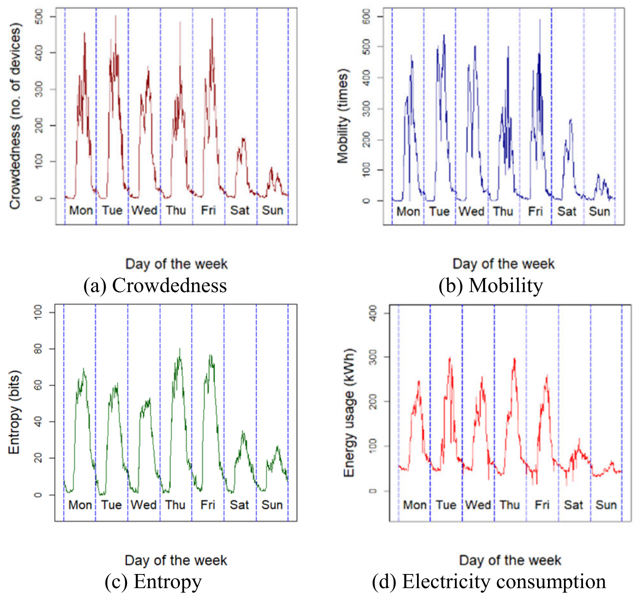
$$H_i(X(q, h, d)) = -\sum\nolimits_{k=1}^{M} P\left(x_k\left(q, h, d\right)\right) \log_2 P\left(x_k(q, h, d)\right), \tag{7}$$

where M is the total number of connections occurred in the data during the quarter q (where $q = 1, 2, 3, 4$) of hour h (where $h = 0, 1, 2, 3, \ldots, 23$) of day d (where $d =$ Monday, Tuesday, Wednesday, …, Sunday). $P\left(x_k(q, h, d)\right)$ denotes the probability of the connectivity being k or $x_k(q, h, d)$, computed as $x_k(q, h, d)/\sum_{k=1}^{M} x_k(q, h, d)$. The connectivity of each AP is quantified by the aggregate count of connections established by any device IDs.

Based on the calculations described previously, Fig. 3 shows the values of crowdedness, mobility, and entropy observed for the building B5 during the week of 13$^{\text{th}}$ – 19$^{\text{th}}$ January 2020, along with its electricity consumption. B5 is a versatile nine-story building, serving as a primary hub for the faculty encompassing the faculty offices, the Department of Computer Engineering, research labs, and lecture rooms. The observed values exhibit distinct patterns, notably differentiated between weekdays and weekends. Crowdedness peaks during daytime hours and diminishes during the night. Interestingly, Wednesday displays lower crowdedness compared to other weekdays, attributed to a reduced number of scheduled lectures on that day. Mobility values, intertwined with crowdedness, offer valuable insights into the dynamics of movement within these crowds. Notably, the observations reflect varying levels of motion corresponding to crowd density. The entropy values exhibit a noteworthy trend, notably lower on Wednesday compared to other weekdays. This discrepancy likely stems from the unique Wednesday schedule with fewer classes and more regular weekly meetings. This predictability renders Wednesdays less random in terms of space utilization compared to other weekdays, aligning with the lower entropy observed.

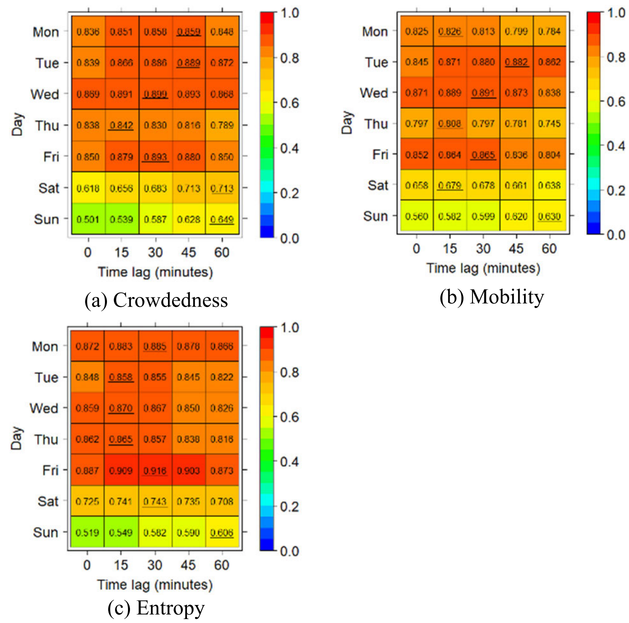
Coefficient of determination or r-squared has been one of the most widely used metrics for interpretation of relationships. So, we employed the coefficient of determination to examine the temporal correlation between the space utilization electricity consumption in the buildings in this study. We examined the resulting r-squared value between the electricity consumption and each of the three space utilization attributes (i.e., crowdedness, mobility, and entropy) with varying time shifts.

For each building, let $y_c = \bar{C}'_i$, $y_m = \bar{M}'_i$, and $y_h = \bar{H}'_i$ denote the set containing the average values of crowdedness,



(a) Crowdedness (b) Mobility (c) Entropy (d) Electricity consumption

FIGURE 3. **Values of (a) crowdedness, (b) mobility, (c) entropy, and (d) electricity consumption of the building B5 during the week of 13th – 19th January 2020.**



(a) Crowdedness (b) Mobility (c) Entropy

FIGURE 4. **The heat map displays the r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for all buildings combined. The underlined number denotes the highest r-squared value observed for each day of the week across different time lags.**

mobility, and entropy, respectively across all APs with the same building ID. Hence, $y_c = \{\overline{c'}(q, h, d)|q \in Quarters, h \in Hours, d \in Days\}$, $y_c = \{\overline{m'}(q, h, d)|q \in Quarters, h \in Hours, d \in Days\}$, and $y_c = \{\bar{H}_i(X(q, h, d))|q \in Quarters, h \in Hours, d \in Days\}$.

Let's consider the r-squared value, denoted as $r_u(t - j)$, representing the correlation between the electricity consumption and one of the space utilization attributes, where $u \in \{y_c, y_m, y_h$. For instance, $r_{y_c}(t - j)$ reflects the r-squared value between electricity consumption and crowdedness j minutes prior. Our investigation covered time lags from 0 to 60 minutes, with a 15-minute interval, constrained by our electricity data. Given multiple data for each day of the week, we calculated the average r-squared value ($\bar{r}_u^d$) to portray the observed correlation for each day. We examined time lags for all three space utilization attributes across all buildings, seeking to identify the maximum values obtained – hence, highest correlation.

III. RESULTS

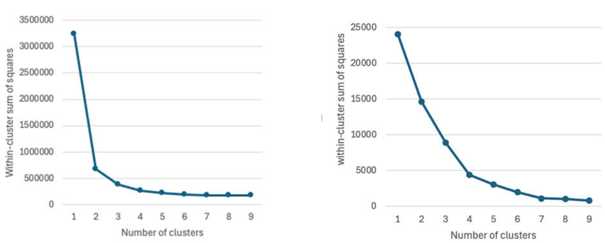
Our analysis began with an examination of the resulting r-squared values obtained from considering all 10 buildings together. The findings are shown in Fig. 4, utilizing a heatmap that portrays the correlation across time lags ranging from 0 to 60 minutes for various days of the week. Notably, the correlation is significantly higher on weekdays compared to weekends. This observation aligns with intuition, given that weekdays are characterized by regular working days and scheduled activities, whereas weekends tend to feature more irregular events. The maximum value observed for each day of the week, considering different time lags, is emphasized by being underlined for clarity.

Crowdedness in a building can have a direct and multi-faceted temporal correlation with electricity consumption due to increased device usage, elevated lighting requirements, higher demand on HVAC systems, frequent elevator usage, shared appliances, audio-visual equipment, and so on. The result in Fig. 4(a) shows that Monday has the highest correlation between crowdedness and electricity consumption occurring at a 45-minute time lag ($r^2 = 0.859$). This suggests that an increase in crowdedness precedes a subsequent rise in electricity consumption by approximately 45 minutes, indicating a conceivable temporal correlation between the two. Fluctuations in crowdedness levels offer valuable insights into the anticipated energy consumption in the near future, particularly within a 45-minute timeframe on Mondays. In fact, the 30-minute time lag also shows a relatively high correlation ($r^2 = 0.858$). From a planning perspective, the 30–45 minute timeframe could serve as an early indicator for anticipated increases in energy usage for Monday. Similar patterns are noticeable on Tuesday, Wednesday, and Friday, where the 30–45 minute timeframe may act as an early indicator of anticipated increases in energy usage. In contrast, Thursday shows a shorter time lag of 15 minutes as the most reliable predictor for energy consumption as far as the crowdedness is concerned.

People's mobility within a building can indeed influence electricity consumption, establishing a causal relationship between the two [22]. From the mobility's perspective, the result in Fig. 4(b) shows that Monday and Thursday both have the highest correlation at the 15-minute time lag, while Wednesday and Friday's mobility correlates most strongly with electricity consumption at a 30-minute time lag. Notably, Tuesday's time lag remains consistent at 45 minutes, mirroring the pattern observed in crowdedness.

Entropy has been used to describe the diversity of individual movement patterns [23], [24]. In the context of the building occupancy, the entropy can serve as a measure of disorder or randomness, which can be applied to analyze the complexity and predictability of occupant movements. This, in turn, can have implications for understanding and predicting electricity consumption in buildings. The result concerning the entropy in Fig. 4(c) reveals that Monday and Friday both exhibit the highest correlation with electricity consumption at a 30-minute time lag. In contrast, Tuesday, Wednesday, and Thursday show that the 15-minute time lag holds the greatest potential as an early indicator of anticipated increases in energy usage within the buildings.

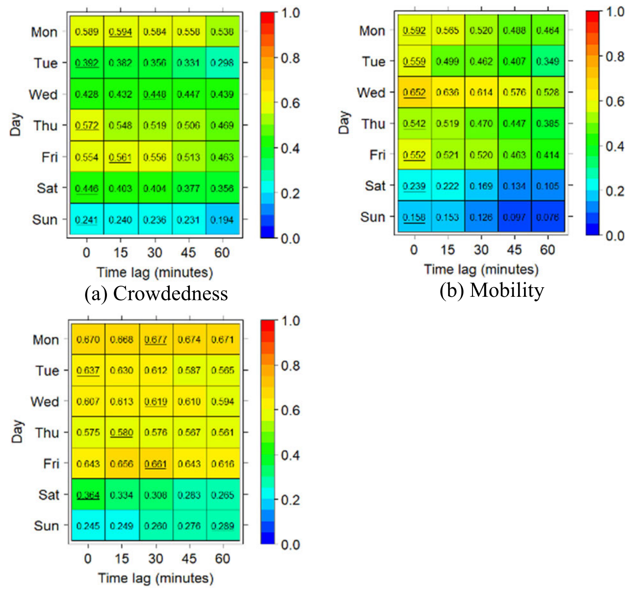
To determine whether the correlation depends on the pattern of electricity consumption across different buildings, we clustered the buildings into groups based on their electricity usage profiles. Each building's energy usage profile is represented by a vector of size $4 \times 24 \times 7 = 672$, capturing the average consumption over four quarters of an hour, 24 hours a day, and seven days a week, which is then used as an input for the clustering. Employing the k-means clustering algorithm [25], in conjunction with the elbow method [26] for selecting the optimal number of clusters (k), resulted in three clusters. The k-means algorithm was chosen for its simplicity, efficiency, and ability to handle large datasets, making it suitable for clustering buildings based on their electricity usage profiles. The elbow method was used to determine the optimal number of clusters by identifying the point where the within-cluster sum of squares (WCSS) starts to level off, ensuring that we do not add unnecessary complexity with too many clusters. The resulting WCSS values are shown in Fig. 5(a) from which k was selected at 3 clusters. This combination of techniques provides a balanced and interpretable clustering result that facilitates our analysis. Cluster 1 comprises B2, B6, and B10, Cluster 2 includes B5, and Cluster 3 consists of B1, B3, B4, B7, B8, and B9.



(a) Electricity usage profiles (b) Space allocation

FIGURE 5. Resulting WCSS values used for consideration in the elbow method for selecting the optimal number of clusters of buildings according to their (a) electricity usage profiles and (b) space allocation.

The correlation results are shown Figs. 6, 7, and 8 for the Clusters 1, 2, and 3, respectively. Overall, it can be observed that Clusters 1 and 3 exhibit a relatively lower correlation



(a) Crowdedness (b) Mobility (c) Entropy

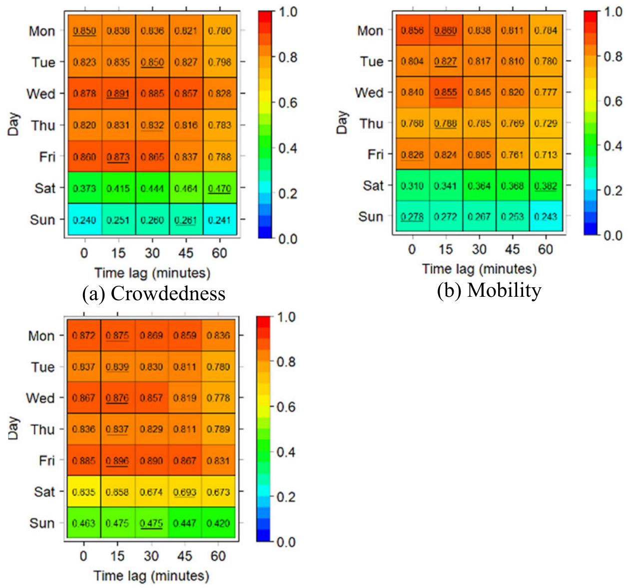
FIGURE 6. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the electricity consumption-based cluster 1's buildings combined.

compared to cluster 2. Notably, Cluster 2 consists only of building B5, a primary hub for the faculty and the most populated high-rise building. The lower correlation in Clusters 1 and 3 may be attributed to the fact that these buildings are low-rise structures situated close to parking spaces, potentially leading to interference from Wi-Fi connections used by nearby commuters outside the buildings. Nevertheless, the entropy emerges as a reliable indicator with a notably high correlation for buildings in the Clusters 2 and 3. The intuitive connection lies in their proximity to parking spaces, whereby entropy serves as a reflection of the fluctuations in people's movement, providing insights into electricity consumption.

Upon grouping buildings according to the allocation of their spaces for various purposes, as outlined in Table 1, we used the k-means clustering algorithm with the elbow method. With its WCSS result shown in Fig. 5(b), the number of optimal clusters was selected at $k = 4$. The outcomes are depicted in Figs. 9, 10, 11, and 12 for Clusters 1, 2, 3, and 4, respectively. Cluster 1 comprises solely building B6, while Cluster 2 encompasses B3, B9, B8, and B10. Cluster 3 is composed solely of B4, and Cluster 4 encompasses B1, B2, B5, and B7.

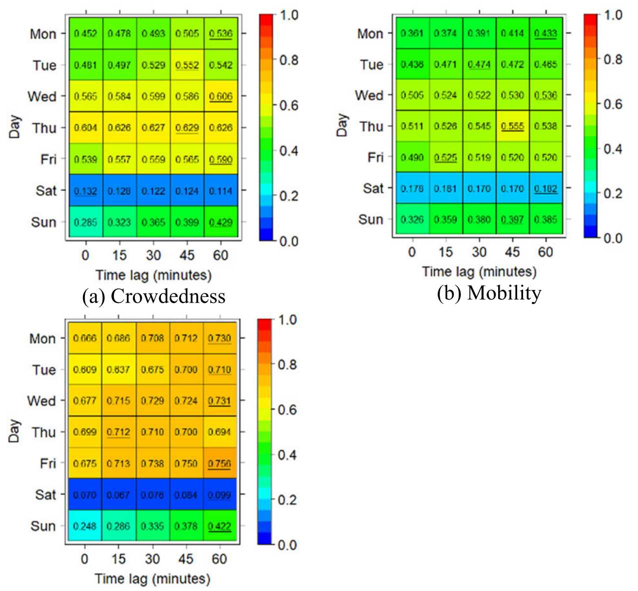
The faculty canteen, denoted as B6 and forming its exclusive Cluster 1, exhibits a relatively weak correlation, likely attributed to the fact that its electricity consumption is not strongly tied to crowd size. Typically, the canteen's lights remain on during business hours, irrespective of the presence of customers. Consequently, fluctuations in people's presence do not significantly impact the electricity consumption of the canteen building, as depicted in the results shown in Figure 8.

The buildings within Cluster 2 demonstrate a notably strong correlation, particularly in terms of entropy.



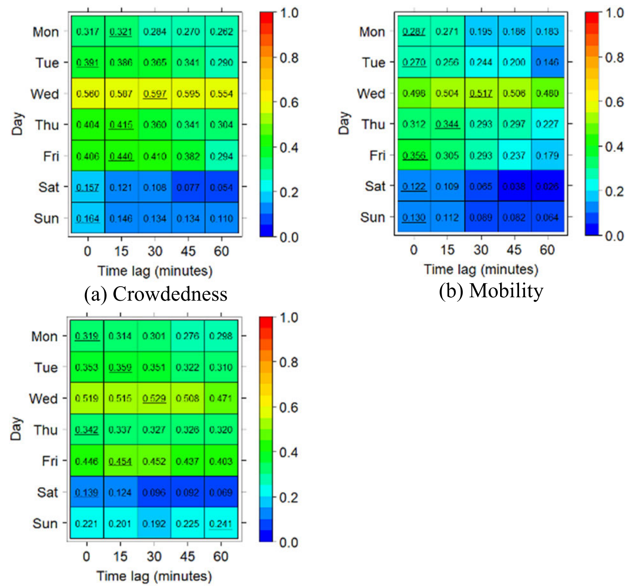
(a) Crowdedness (b) Mobility (c) Entropy

FIGURE 7. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the electricity consumption-based cluster 2's building.



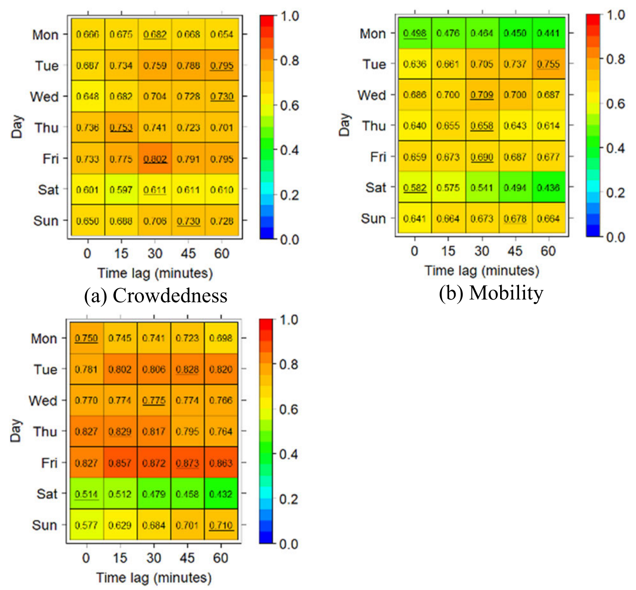
(a) Crowdedness (b) Mobility (c) Entropy

FIGURE 8. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the electricity consumption-based cluster 3's buildings combined.



(a) Crowdedness (b) Mobility (c) Entropy

FIGURE 9. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the space allocation-based cluster 1's building.



(a) Crowdedness (b) Mobility (c) Entropy

FIGURE 10. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the space allocation-based cluster 2's buildings.

These buildings are designated for use by individuals from specific departments and adhere to regular schedules for area utilization, such as lectures and meetings. For instance, B3 is exclusively utilized by individuals from the Department of Electrical Engineering's high voltage discipline, while B10 serves solely the field of survey engineering. The specificity of usage patterns in these buildings establishes a correlation that aligns well with electricity consumption, enabling us to discern their robust temporal relationship.

Cluster 3 comprises solely building B4, the primary office facility for Mechanical Engineering (ME), housing staff offices exclusively, without any lecture rooms. Interestingly, the observed correlation within this cluster is relatively low. This might be attributed to a diminished office occupancy, particularly among academic staff who, during non-teaching hours, often deliver lectures in other buildings and engage in research labs located elsewhere. Consequently, ME academic staff may not consistently utilize their offices,

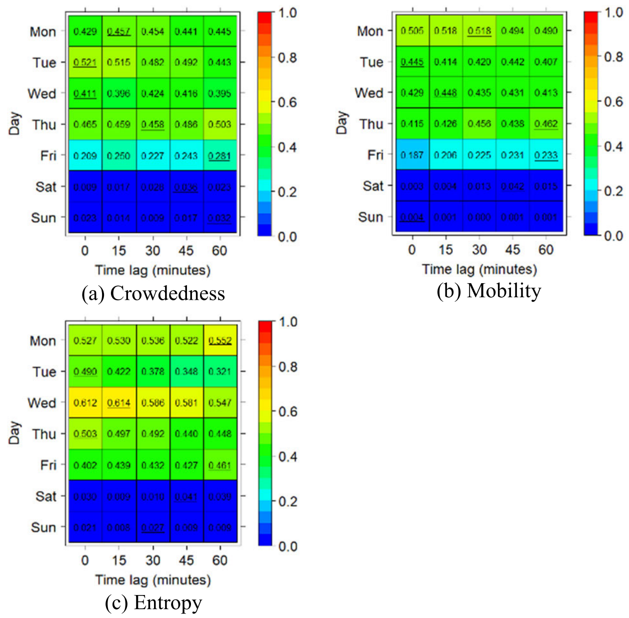


FIGURE 11. **The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the space allocation-based cluster 3's buildings combined.**

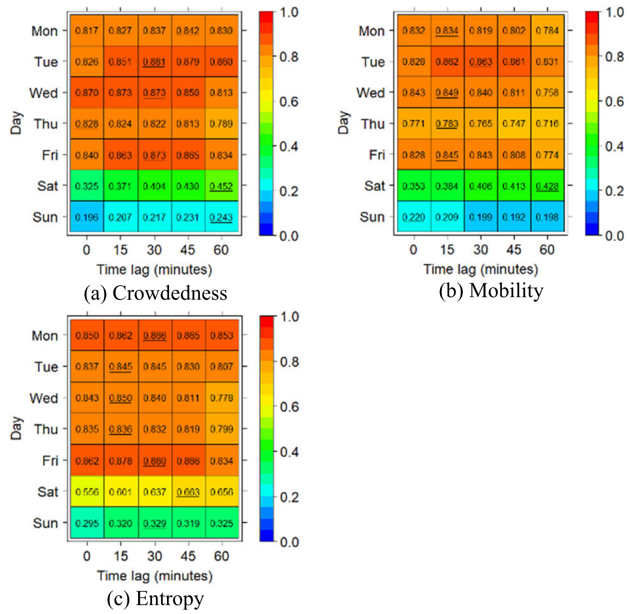


FIGURE 12. **The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the space allocation-based cluster 4's buildings combined.**

while the building's electricity consumption is driven by supporting staff who regularly occupy the premises. Furthermore, with no lectures conducted in this building, student occupancy and movement are relatively low. The limited variability in both building occupancy and electricity consumption values may contribute to an insignificant correlation.

Cluster 4's buildings show a strong correlation. Since all buildings have rooms allocated for lectures, it creates dynamics in space occupancy. Typically, lights and air conditioners are turned on only while the lecture room is occupied. So, the room occupancy and the movement of people within the buildings can certainly ripple the use of electricity, causing a high correlation at different time lags.

Cluster 4 demonstrates a pronounced correlation among its buildings. This strong correlation is attributed to the presence of lecture rooms in each building, introducing dynamic changes in space occupancy. Typically, the activation of lights and air conditioners aligns with the occupancy of these lecture rooms. As a result, the movement of individuals within the buildings, particularly during lecture sessions, markedly influences electricity consumption, leading to a discernible temporal relationship across various time lags.

The examination was extended to individual buildings using a similar approach. However, the observed correlations were not statistically significant, with the exception of building B5, the primary hub of the faculty. The significant characteristics of B5, such as its high-rise structure, expansive area, and elevated occupancy rate, presumably are key factors in capturing a meaningful temporal relationship between space utilization and electricity consumption. Detailed results from the examination of individual buildings are provided in the Appendix (Fig. 13). For practical insights into energy management, it is important to note that the faculty's Facilities Management Office estimates the ratio of HVAC to lighting and other electricity usage to be approximately 70:30%. This significant proportion of HVAC usage highlights the need to prioritize HVAC system optimization in our energy management strategies.

In addition to space utilization, other potential factors can influence a building's electricity consumption. One of the main factors is weather conditions, including temperature and humidity. Research has shown that external weather conditions can impact energy consumption in residentials and buildings [27], [28]. To account for this, we gathered weather condition data, specifically temperature and humidity as the most significant factors affecting thermal comfort [29], during the study period (January 9 - February 3, 2020) from a land-based weather station located on the top of Building B5 in our study area. The observed data is shown alongside electricity usage in Fig. 14 (in the Appendix). To assess the impact of weather conditions on electricity usage, we examined the correlation between these variables. The results, presented in Fig. 15 (Appendix), show a relatively low correlation, with a correlation of 0.4789 between temperature and electricity usage, and 0.3289 between humidity and electricity usage. This suggests that weather conditions do not play as significant a role as the space utilization attributes explored in this study. Regarding room temperature and humidity, our interviews with the engineers responsible for the faculty's buildings and regular users indicate that the indoor temperature typically ranges from 24-27 degrees Celsius, and the humidity ranges from 55-60%.

IV. CONCLUSION

In conclusion, this paper presents initial exploration into understanding the temporal correlation between space utilization and electricity consumption within buildings, addressing a significant gap in current research. Leveraging Wi-Fi probe data for physical space segmentation, the study introduces a novel approach for comprehending space utilization, using Wi-Fi connectivity data from 10 academic buildings at the Faculty of Engineering, Chiang Mai University, Thailand. By examining factors such as crowdedness, mobility, and entropy inferred from the Wi-Fi data, and correlating these with changes in electricity consumption, the research aims to uncover strategies for optimizing energy use in real-time, positioning itself at the forefront of discussions on sustainable building practices and energy management.

However, it is crucial to acknowledge several limitations inherent in the study. Firstly, while the chosen features provide valuable insights, there may be other potentially influential features within the Wi-Fi data characterizing area usage that were not explored in this research, opening avenues for future investigations. Additionally, network connection issues may have led to connecting and disconnecting events in the logs, potentially impacting the analysis, particularly in terms of mobility and entropy calculations. Although these issues were deemed marginal, they warrant consideration in future investigations.

Furthermore, the absence of ground truth confirmation for the resulting space utilization inference raises questions about the accurate measurement of area utilization. Approaches and methods for sensing and assessing space utilization are open questions that could be addressed in future work.

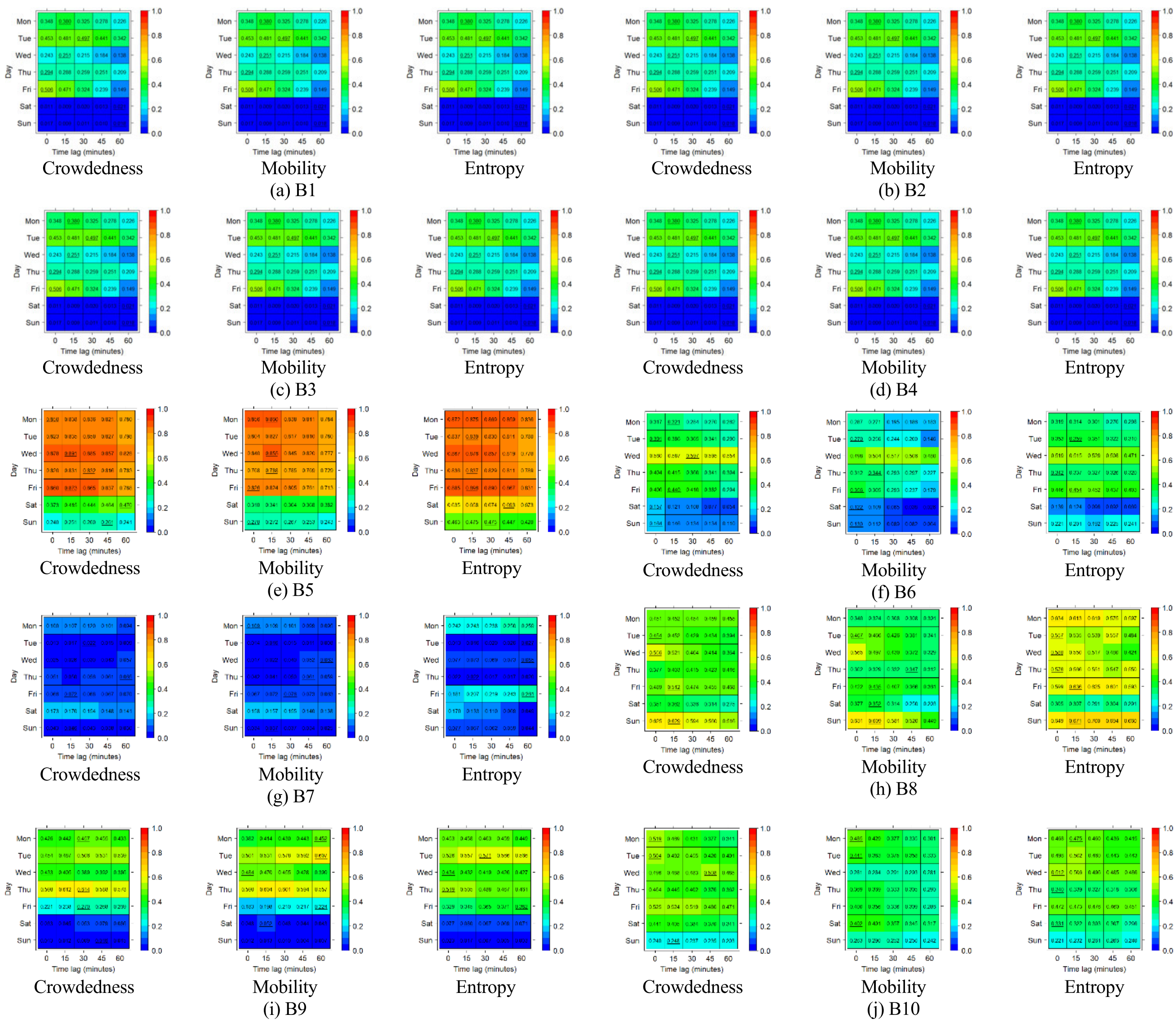
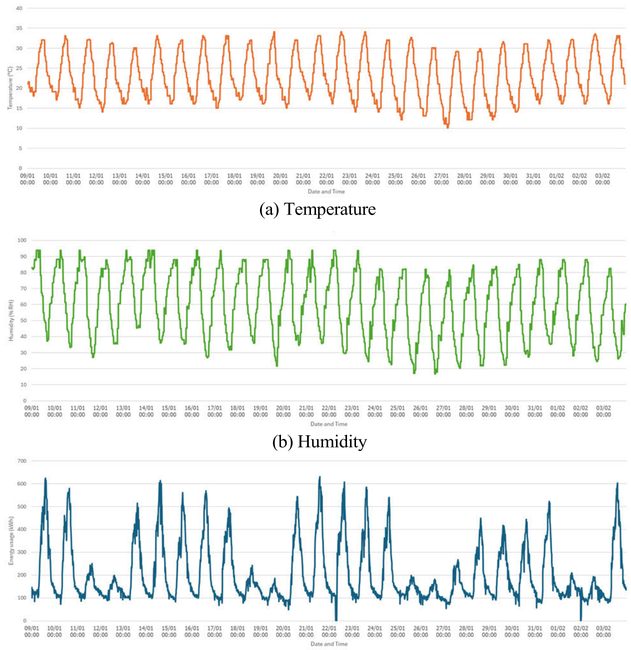


FIGURE 13. The r-squared values correlating electricity consumption with (a) crowdedness, (b) mobility, and (c) entropy across various time lags for the space allocation-based cluster 4's buildings combined.

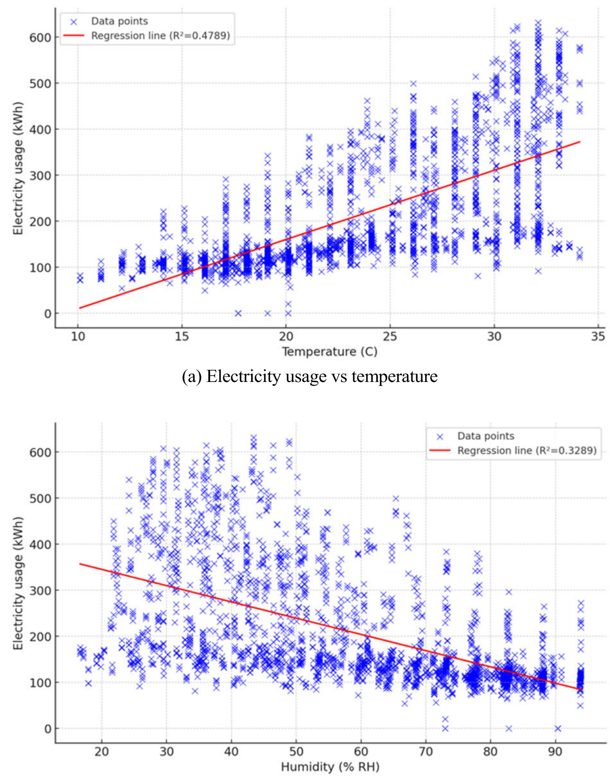


(a) Temperature

(b) Humidity

(c) Electricity usage

FIGURE 14. **Weather condition data collected from a land-based weather station during the study period: (a) temperature and (b) humidity, along with the total electricity consumption from all buildings.**



(a) Electricity usage vs temperature

(b) Electricity usage vs humidity

FIGURE 15. **Scatter plots showing relatively low correlation between electricity usage and weather conditions: (a) temperature ($r^2 = 0.4789$) and (b) humidity ($r^2 = 0.3289$).**

The sampling rate, constrained by the electricity data, was set at a 15-minute interval, limiting the granularity of the analysis. A shorter interval could offer a more detailed examination of the temporal relationship. Finally, the study was conducted over regular days, and the temporal relationship may be influenced by special schedules, holidays, and social events. Other unexplored factors such as building age and surrounding conditions present additional avenues for further exploration. Despite these limitations, this study contributes valuable insights into the intricate dynamics between space utilization and electricity consumption, laying the groundwork for future research in sustainable building practices and energy management.

It is important to note that the coefficient of determination (r^2) was used in this study to quantify the strength of the relationship between space utilization and electricity consumption. However, we acknowledge that correlation does not imply causation. To determine causation, more advanced methods such as Bayesian networks or Structural Causal Models (SCMs) should be considered. Future research could leverage these methods to further investigate the causal relationships between space utilization and electricity consumption in buildings.

APPENDIX

Here are the detailed results from the examination of individual buildings.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Information Technology Service Center (ITSC), Chiang Mai University, for providing Wi-Fi probe data, and the Energy Research and Development Institute-Nakornping, Chiang Mai University, for supplying electricity consumption data.

REFERENCES

[1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy Buildings*, vol. 128, pp. 198–213, Sep. 2016, doi: 10.1016/j.enbuild.2016.06.089.

[2] C. Eon, J. K. Breadsell, G. M. Morrison, and J. Byrne, "The home as a system of practice and its implications for energy and water metabolism," *Sustain. Prod. Consumption*, vol. 13, pp. 48–59, Jan. 2018, doi: 10.1016/j.spc.2017.12.001.

[3] J. Torriti, "Understanding the timing of energy demand through time use data: Time of the day dependence of social practices," *Energy Res. Social Sci.*, vol. 25, pp. 37–47, Mar. 2017, doi: 10.1016/j.erss.2016.12.004.

[4] M. Bourdeau, X. Guo, and E. Nefzaoui, "Buildings energy consumption generation gap: A post-occupancy assessment in a case study of three higher education buildings," *Energy Buildings*, vol. 159, pp. 600–611, Jan. 2018, doi: 10.1016/j.enbuild.2017.11.062.

[5] I. Qaisar, K. Sun, Q. Zhao, T. Xing, and H. Yan, "Multi-sensor-based occupancy prediction in a multi-zone office building with transformer," *Buildings*, vol. 13, no. 8, p. 2002, Aug. 2023, doi: 10.3390/buildings13082002.

[6] B. Dong, V. Prakash, F. Feng, and Z. O'Neill, "A review of smart building sensing system for better indoor environment control," *Energy Buildings*, vol. 199, pp. 29–46, Sep. 2019, doi: 10.1016/j.enbuild.2019.06.025.

[7] W. Alsafery, O. Rana, and C. Perera, "Sensing within smart buildings: A survey," *ACM Comput. Surveys*, vol. 55, no. 13, pp. 1–35, Dec. 2023, doi: 10.1145/3596600.

[8] V. L. Erickson, M. Á. Carreira-Perpiñán, and A. E. Cerpa, "Occupancy modeling and prediction for building energy management," *ACM Trans. Sensor Netw.*, vol. 10, no. 3, pp. 1–28, Apr. 2014, doi: 10.1145/2594771.

[9] H. Latha, S. Patil, and P. G. Kini, "Influence of architectural space layout and building perimeter on the energy performance of buildings: A systematic literature review," *Int. J. Energy Environ. Eng.*, vol. 14, no. 3, pp. 431–474, Sep. 2023, doi: 10.1007/s40095-022-00522-4.

[10] P. W. Tien, S. Wei, J. K. Calautit, J. Darkwa, and C. Wood, "Real-time monitoring of occupancy activities and window opening within buildings using an integrated deep learning-based approach for reducing energy demand," *Appl. Energy*, vol. 308, Feb. 2022, Art. no. 118336, doi: 10.1016/j.apenergy.2021.118336.

[11] X. Gui, Z. Gou, and F. Zhang, "The relationship between energy use and space use of higher educational buildings in subtropical Australia," *Energy Buildings*, vol. 211, Mar. 2020, Art. no. 109799, doi: 10.1016/j.enbuild.2020.109799.

[12] S. Chaikul, S. Phithakkitnukoon, and C. Ratti, "Xplaces: Segmenting physical space through Wi-Fi traces using eigendecomposition and X-means," *IEEE Access*, vol. 10, pp. 17377–17388, 2022, doi: 10.1109/ACCESS.2022.3150006.

[13] C. Shen, "Analysis of detrended time-lagged cross-correlation between two nonstationary time series," *Phys. Lett. A*, vol. 379, no. 7, pp. 680–687, Mar. 2015, doi: 10.1016/j.physleta.2014.12.036.

[14] W. N. Probst, V. Stelzenmuller, and H. O. Fock, "Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: An exemplary analysis of North Sea fish population indicators," *ICES J. Mar. Sci.*, vol. 69, no. 4, pp. 670–681, May 2012, doi: 10.1093/icesjms/fss015.

[15] H. Liu, Y. Li, C. Liu, G. Shen, and H. Xiang, "Pavement distress initiation prediction by time-lag analysis and logistic regression," *Appl. Sci.*, vol. 12, no. 22, p. 11855, Nov. 2022, doi: 10.3390/app122211855.

[16] C. Smith, A. Dobroshynskyi, and S. McIntosh, "Quantifying local energy demand through pollution analysis," in *Machine Learning, Optimization, and Data Science*. Cham, Switzerland: Springer, 2020, pp. 504–515.

[17] K. Qin, Y. Xu, C. Kang, S. Sobolevsky, and M.-P. Kwan, "Modeling spatio-temporal evolution of urban crowd flows," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 12, p. 570, Dec. 2019, doi: 10.3390/ijgi8120570.

[18] A. Binthaisong, J. Srichan, and S. Phithakkitnukoon, "Wi-crowd: Sensing and visualizing crowd on campus using Wi-Fi access point data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2017, pp. 441–447, doi: 10.1145/3123024.3124413.

[19] R. Smarzaro, C. A. Davis, and J. A. Quintanilha, "Creation of a multimodal urban transportation network through spatial data integration from authoritative and crowdsourced data," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 7, p. 470, Jul. 2021, doi: 10.3390/ijgi10070470.

[20] E. Graells-Garrido, F. Serra-Burriel, F. Rowe, F. M. Cucchietti, and P. Reyes, "A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in Barcelona," *PLoS One*, vol. 16, no. 5, May 2021, Art. no. e0250080, doi: 10.1371/journal.pone.0250080.

[21] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[22] Z. Huang and Z. Gou, "Electricity consumption variation of public buildings in response to COVID-19 restriction and easing policies: A case study in Scotland, U.K.," *Energy Buildings*, vol. 267, Jul. 2022, Art. no. 112149, doi: 10.1016/j.enbuild.2022.112149.

[23] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010, doi: 10.1126/science.1177170.

[24] D. D. C. Teixeira, J. M. Almeida, and A. C. Viana, "On estimating the predictability of human mobility: The role of routine," *EPJ Data Sci.*, vol. 10, no. 1, pp. 1–30, Dec. 2021, doi: 10.1140/epjds/s13688-021-00304-8.

[25] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci.*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.

[26] B. Purnima and K. Arvind, "EBK-Means: A clustering technique based on elbow method and K-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 1–27, 2014.

[27] M. G. Fikru and L. Gautier, "The impact of weather variation on energy consumption in residential houses," *Appl. Energy*, vol. 144, pp. 19–30, Apr. 2015, doi: 10.1016/j.apenergy.2015.01.040.

[28] M. R. Sanzana, M. O. M. Abdulrazic, J. Y. Wong, T. Maul, and C.-C. Yip, "Effects of external weather on the water consumption of thermal-energy-storage air-conditioning system," *Energy Nexus*, vol. 10, Jun. 2023, Art. no. 100187, doi: 10.1016/j.nexus.2023.100187.

[29] S. Zhang and Z. Lin, "Effective moisture temperature: Ventilation performance index accounting for effects of air temperature and relative humidity on thermal comfort," *Building Environ.*, vol. 243, Sep. 2023, Art. no. 110625, doi: 10.1016/j.buildenv.2023.110625.

SUNSIKA CHAIKUL received the B.S. degree in computer science from the Nation University and the M.S. degree in computer engineering from Chiang Mai University, Thailand, where she is currently pursuing the Ph.D. degree with the Department of Computer Engineering. Her research interests include urban data science and visual analytics.

YOTTANA KHUNATORN (Member, IEEE) received the B.Eng. degree in mechanical engineering from Chiang Mai University, Thailand, in 1992, and the M.S. and Ph.D. degrees in mechanical engineering from the University of Colorado, in 1996 and 2002, respectively. He was the Deputy Director with the Energy Research and Development Institute-Nakornping, Chiang Mai University, from March 2011 to 2023. He is currently an Assistant Professor with the Department of Mechanical Engineering, Faculty of Engineering, Chiang Mai University. His research is on cardiovascular engineering, near zero energy building, and energy forecasting and optimization. He is a Senior Advisor of the IEEE Thailand Section SIGHT (Special Interest Group on Humanitarian Technology).

SANTI PHITHAKKITNUKOON received the B.S. and M.S. degrees in electrical engineering from Southern Methodist University, Dallas, USA, in 2003 and 2005, respectively, and the Ph.D. degree in computer science and engineering from the University of North Texas, USA. He is currently a Full Professor of data engineering (pending official royal appointment for full professorship) and the Head of the Computer Engineering Department, Faculty of Engineering, Chiang Mai University, Thailand. Before joining Chiang Mai University, he was a Lecturer in computing with The Open University, U.K.; a Research Associate with Newcastle University, U.K.; and a Postdoctoral Fellow with the SENSEable City Laboratory, Massachusetts Institute of Technology, USA. His research focuses on urban informatics.