# Floor-level Occupancy Estimation of a Multi-Story Building Using Coarse Wi-Fi Data

by

Ryan Bobo, B.S.

Directed research submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Applied Geography
with a Major in Geographic Information Science
May 2023

Committee Members:

Dr. T. Edwin Chow

Dr. Yihong Yuan

## DEDICATION

To my son, Lennox.

## ACKNOWLEDGEMENTS

I want to express my deepest gratitude to Dr. Edwin Chow, my advisor, for his relentless patience, direction, and feedback. I could only have undertaken this journey with my professor and committee member, Dr. Yihong Yuan, whose classes and labs prepared me for this endeavor. I am deeply indebted to Dr. Jennifer Jensen and Dr. David Parr for so expertly introducing me to the geospatial technologies that led me down this path. Additionally, I would like to thank Michael O'Connor for planting the seed that became this research and advising along the way.

I am grateful to my family and friends for the love and encouragement that kept my spirits and motivation high during this process. This undertaking would not have been possible without my wife, Drew. Thank you for your patience, love, and support.

## TABLE OF CONTENTS

**Page**

## LIST OF TABLES

**Table** **Page**

## LIST OF FIGURES

**Figure** **Page**

## LIST OF ABBREVIATIONS

(In order of appearance)

| Abbreviation | Description |
|---|---|
| AP | Access Point |
| MAC | Media Access Control |
| RSS | Received Signal Strength |
| Wi-Fi | Wireless Fidelity |
| RF | Radio Frequency |
| RFID | Radio Frequency Identification |
| GPS | Global Positioning System |
| RPC | Resident Population Classification |
| IT | Information Technology |
| MED | Medium |
| STF | Stay-Time Filter |
| IFT | Initial Floor Filter |
| SFT | Succeeding Floor Threshold |
| IFF | Isolated Floor Filter |

## ABSTRACT

In recent years, there has been an extraordinary increase in wireless capable devices and network infrastructure, which spawned a corresponding rise in data produced from the interactions of these technologies. Mobile devices constantly roam ,
leading to a perpetual dialog between a mobile device and wireless access points. This dialogue generates a continuous stream of device-specific data, including but not limited to a device's media access control address, time of access, and received signal strength. Given the knowledge of the access point's location and received signal strength, it is possible to infer the position of user devices and estimate their mobility and occupancy. This paper presents two methods for accurately measuring floor-level occupancy in a multi-story building at Texas State University using coarse Wi-Fi log data. The first method employs a static filter, while the second incorporates user-role data and user location to create a dynamic filter. Quantitative methods are used to evaluate these filters against field-collected reference data and existing internal people-counting sensors. Our results demonstrate that the dynamic filter, leveraging variable thresholds, provides a more accurate estimation of occupancy compared to the fixed 5-minute static filter which consistently overestimated occupancy. This research sheds light on the potential of dynamic filters derived from user-role data for precise floor-level occupancy estimations, with implications for various applications

# 1. INTRODUCTION

In recent years, there has been an extraordinary increase in wireless capable devices and network infrastructure, which spawned a corresponding rise in data produced from the interactions of these technologies. Mobile devices constantly roam, assuming the network settings are activated, leading to a perpetual dialog between a mobile device and wireless access points (APs). This dialogue generates a continuous stream of device-specific data, including but not limited to a user's media access control (MAC) address, time of access, and received signal strength. Given the knowledge of the access point’s location and received signal strength (RSS), it is possible to infer the position of user devices, as well as estimate their mobility and occupancy (Uras, Cossu and Atzori 2019, Sapiezynski et al. 2015).

In building management, knowing the location and quantity of people is incredibly important for understanding individual and group interactions in an indoor environment. Occupancy estimation is used in applications such as workplace analysis (Regodón et al. 2021), emergency management (Khoche et al. 2021), and building energy estimation (Rafsanjani and Ghahramani 2019), to name a few. Although, until recently, much of the data for this research was made possible through field surveys that rely on humans to manually count individuals (Traunmueller et al. 2018). These methods are accurate and often used to validate alternative approaches of occupancy estimation, but they come with an inherently large human and hardware cost. New methods using cameras, environmental sensors, and radio frequency (RF) devices have been explored in response.

Monitoring building occupancy with wireless fidelity (Wi-Fi) data increases the size and scope of what is possible compared to traditional practices. Additionally, it offers several benefits compared to image processing and environmental sensors. Firstly, given the widespread use of wireless technology, wireless infrastructure is widely available in most buildings nowadays (Wang et al. 2019b, Chen and Ahn 2014). This approach minimizes implementation obstacles and additional hardware costs for contemporary facilities. Furthermore, Wi-Fi occupancy estimation is much less computationally intensive than image processing (Fayed et al. 2022).

To estimate occupancy within an area using Wi-Fi data, it is first essential to infer the device user's position. Numerous techniques are used to calculate a user's indoor position using wireless technology, each of which implies varying accuracy and spatial resolution. Common spatial resolutions of occupancy estimation studies reported in the literature include floor-level (Anand et al. 2021), zone-level (Regodón et al. 2021), room-level (Jia, Srinivasan and Raheem 2017), and individual-level (Petrenko et al. 2014). Fingerprinting is a seminal indoor positioning method used in the vast majority of wireless occupancy studies and has the potential to achieve individual accuracies of 1-3 meters (Bi et al. 2021). Despite this relatively high level of accuracy, many works that utilize indoor positioning techniques are diluted by aggregating positions into floor-level (Anand et al. 2021) or indoor zones (Mashuk et al. 2021, Regodón et al. 2021, Khoche et al. 2021) when performing an occupancy analysis.

Given that most wireless occupancy estimation is accomplished using highly accurate indoor positioning systems, there is a lack of literature using coarser Wi-Fi log data. To fill this gap, this research aims to establish floor-level occupancy of a multi-story academic building at Texas State University using existing wireless infrastructure and coarse Wi-Fi log data. The Wi-

Fi data used in this case is defined as coarse because each log record is limited to the single AP a device is connected to (as opposed to multiple APs), which prohibits highly accurate positioning using modern indoor positioning techniques. This study will validate the occupancy estimation with a conventional field-based occupancy count and an internal radio frequency identification (RFID) occupancy sensor. Although the data has limited potential in terms of accuracy, it is unique with the inclusion of the user's role in each record. This allows for the distinction between student, staff, and guest to be used to potentially improve the occupation estimation model. Therefore, the lack of multiple APs and the presence of user-roles establishes three research questions as follow.

1. Is there any significant difference between floor-level occupancy estimation with or without user-role data?
2. Does user-role data improve the accuracy of floor-level estimated occupancy when compared to ground truth measurements?
3. Is there any significant difference between floor-level occupancy estimation when compared to an existing internal RFID occupancy sensor?

The rest of this paper is organized as follows: Section two will introduce literature surrounding indoor positioning and its contribution to occupancy estimations. Section three will more completely describe the data and methodology used to estimate occupancy. Sections four and five will include results and a discussion and Section six will offer concluding statements and limitations.

## 2. LITERATURE REVIEW

Occupancy studies utilizing Wi-Fi data generally fall into three types, including indoor (Regodón et al. 2021), outdoor (Traunmueller et al. 2018), and hybrid approaches (Prentow et al. 2015). In terms of geographic scales, indoor studies can generally be further divided into floor-level (Anand et al. 2021), zone-level (Regodón et al. 2021), room-level (Jia et al. 2017), and individual-level (Petrenko et al. 2014). This literature review consists of four sections that examine 1) Wi-Fi based occupancy estimations, 2) Indoor positioning systems, 3) the use of user-role information, and 4) research questions.

### *2.1. WI-FI-BASED OCCUPANCY ESTIMATION*

Early works employing wireless network data largely concern network activity and performance (Kotz and Essien 2002, Schwab and Bunt 2004, Kim and Kotz 2005). With the common goal of producing quality information for managing a large-scale network, these studies focus on the spatial distribution of network traffic and associated mobility. As the popularity and availability of mobile devices became more prevalent, more human-centric works began to appear in the literature. For example, a model for measuring mobility uses the arrivals and departures measured at APs (Kim and Kotz 2005). By counting and aggregating the number of visits an AP received by the hour, it was possible to monitor the temporal distribution of AP clusters, e.g., peak traffic hours. This allowed the discovery of varying diurnal patterns of users on the campus (Kim and Kotz 2005). As the field of indoor wireless positioning evolved, advanced methodologies from Global Positioning System (GPS) were applied to Wi-Fi traces to increase accuracy and bolster insights into applications such as indoor navigation (Chai et al. 2012), pedestrian dynamics (Danalet, Michel and Farooq 2012), and mobility patterns (Petrenko et al. 2014).

Wi-Fi log data is a steady stream of information conveying the interactions between all network-capable devices and connected network equipment. The connected network device determines the Wi-Fi log data frequency and can fluctuate based on device type and age, power-saving functions, and whether or not the device is active (Wang, Tse and Chan 2019a). To get a sense of scale, hundreds of millions of Wi-Fi logs were collected over 16 months by 19 APs strategically placed by Wang, Zhu and Miao (2016) on a college campus. Given the sheer magnitude of this data, threshold filters are often used to screen out unwanted devices and are essential for establishing occupant counts. This filtering process is primarily designed to distinguish mobile users ("short connections") and static devices ("long connections") such as desktops and printers (Min et al. 2021). Static devices are found in most large-scale networks and can be defined as "non-human" devices that are continuously recorded and remain immobile (Wang et al. 2019a). For example, Min et al. (2021) utilized connection thresholds of less than 5-minutes to identify mobile users and greater than 12 hours for static devices. An alternative approach is introduced by Ciftler et al. (2018) where a statistical method is utilized to evaluate the distribution of MAC addresses per quantity of probe requests. This filters out unwanted users, which amounted to around 60% of the total dataset (Ciftler et al. 2018). This method operates off the assumption that static devices will send the most significant amount of wireless log entries, and mobile users will have the least.

Collecting "ground truth" data is necessary for determining the accuracy of Wi-Fi occupancy estimations. In Wang and Shao (2017), the authors performed multiple manual occupancy samples in a non-intrusive walk-through of the study areas. Sign-in sheets have also been implemented in smaller areas (Regodón et al. 2021) where employees of a shared office were requested to sign with time information each time they occupied a workstation for more than 5-minutes (Regodón et al. 2021). Moreover, cameras can also be employed to count occupants manually (Wang et al. 2019a) or by image processing techniques (Tang et al. 2020).

### *2.2. INDOOR POSITIONING*

Modern wireless occupancy estimation hinges on indoor positioning techniques, such as fingerprinting (Xia et al. 2017), triangulation (Wang and Shao 2017), and trilateration (Xia et al. 2017). These methods provide varying levels of accuracy and require RSS measurements from multiple APs (Elgwad, Ashry and Sheta 2019). Fingerprinting is considered one of the most accessible (Elgwad et al. 2019) and utilized (Xia et al. 2017) indoor positioning methods (Elgwad et al. 2019). This is mainly because it doesn't require knowledge of the AP's locations or "line-of-sight" to APs, and is not affected by signal loss from obstacles (Elgwad et al. 2019). Fingerprinting has been combined with additional methods in an attempt to increase accuracy (Bi et al. 2021, Ciftler et al. 2018, Lee, Jung and Han 2021, Ravi and Misra 2020, Wang et al. 2019a), but in its purest form, it relies on two main steps. Firstly, a "radio map" is created by collecting RSS measurements of all available APs for a given area of interest, then locations are assigned to each collection of measurements (Elgwad et al. 2019). Next, positioning is derived by comparing network device measurements of three or more APs to the established radio map (Elgwad et al. 2019). Fingerprinting is the most widely used indoor positioning method due to its low setup time and relatively high accuracy (Elgwad et al. 2019). Regodón et al. (2021) applied fingerprinting techniques by resampling positions based on 3-minute windows and assigning the positions into zones to calculate occupancy.

Like fingerprinting, triangulation, and trilateration require information from multiple APs. Although, they have the additional requirement of establishing the locations of each AP, which can be time-consuming and costly compared to fingerprinting. A triangulation algorithm is applied by Wang and Shao (2017) to examine occupancy and behavior patterns in a multiuse campus building. Indoor positioning methods that employ triangulation rely on angles from at least two APs or "ground-control stations" to establish positioning (Elgwad et al. 2019, Wang and Shao 2017). Additional drawbacks of triangulation include relatively low accuracies and high hardware costs because most standard network equipment cannot measure angles (Elgwad et al. 2019, Wang and Shao 2017). Instead of angles, trilateration uses known distances from RSS values of three or more APs instead of angles to calculate a network device's position (Biczók et al. 2014). This makes trilateration more accessible than triangulation when using existing network hardware.  WazeMap, an indoor navigation application, utilized trilateration to measure mobility and building utilization and received relatively coarse accuracies ranging from "5-10 meters" (Biczók et al. 2014). The authors extracted location searches, routes, and periodic geographic position logs from the application to analyze movement across the extent of a college campus (Biczók et al. 2014). Both trilateration and triangulation require knowledge of AP location. Combined with relatively low accuracies compared to fingerprinting, this made them less practical and, therefore, less represented in the literature.

Despite the heavy use of accurate indoor positioning methods, many works estimating occupancy resample their positions into spatial zones (Jia et al. 2017, Ravi and Misra 2020, Wang et al. 2019a). Assigning occupancy measurements to zones is prominent in occupancy estimation because it enables further analysis and visualization of "time-series data" (Chen and Ahn 2014). There are two distinct methods for delineating indoor zones found in the literature. The first uses an equal-sized grid overlain on a floorplan (Ciftler et al. 2018). Alternatively, zones have also been constructed based on logical divides of usable space. For instance, (Regodón et al. 2021) delineated zones in an office environment based on groups of workstations. Moreover, (Wang et al. 2019a) introduced a method that utilized one zone that accounted for an entire teaching theatre to establish a measurement of attendance.

A unique challenge of this work is that the Wi-Fi logs are limited to the AP a user is connected to (i.e. one AP per record). Due to this data constraint, standard indoor positioning methods that require multiple APs, such as fingerprinting, triangulation, and trilateration, are not applicable. An additional distinction of the data used in this work is that the RSS values are only accounted for during timeout and roaming network actions. According to network professionals at Texas State University, the coarse nature of the available data is due to the configuration and interactions between the wireless controller and the software used to access the data (O'Connor 2021). In this work, the indoor positioning methods mentioned above are only possible with the presence of multiple APs in the Wi-Fi log data. An additional limitation is found in roaming devices that prevent zone-level occupancy estimations. Once a mobile device is connected to an AP, it will stick to it (stay connected) until the RSS value reaches less than or equal to -75 dBm. This lack of precision creates difficulties when developing occupancy zones smaller than floor-level. Due to similar limitations, Anand et al. (2021) performed an energy consumption model limited to floor-level occupancy estimations. In their work, additional steps were taken to classify user-roles based on daily schedules, which allowed particular roles to be assigned to more distinct zones (Anand et al. 2021). Based on ground truth measurements collected by the researchers, this method yielded an average zonal accuracy of 87% (Anand et al. 2021). Further conversation regarding accuracy limitations will be discussed in more detail in section 3.3.

### *2.3. USER-ROLE*

Work has been done to distinguish the roles of users based on their spatiotemporal patterns (Anand et al. 2021), behavioral classifications (Ruiz-Ruiz et al. 2014, Prentow et al.

2015) and semantic trajectories (Wang et al. 2016), but there appears to be no attempt in the literature to use pre-existing role data to facilitate estimations of occupancy. existing work can delineated classifications of patients, staff, and visitors in a hospital complex based on knowledge obtained by hospital professionals (Prentow et al. 2015, Ruiz-Ruiz et al. 2014). For instance, hospital staff members could be distinguished and classified based on their high mobility and consistent spatiotemporal patterns across the hospital complex. Furthermore, users assigned to a patient role were identified based on their limited movement and facility access (Prentow et al. 2015). The ability to estimate users' roles and spatiotemporal patterns allows hospital administrators to make more effective facility planning decisions. In a university campus setting, Wang et al. (2016) leverage Wi-Fi trace data collected from 19 Wi-Fi devices over six months to measure the similarity of trajectories and make inferences about user relationships. The authors introduce the idea of a "Resident Population Classification" (RPC), which extracts users that reside in a particular building based on a typical living schedule. The RPC was created by further categorizing buildings into the following groups: "teaching building, canteen, laboratory building and dormitory" (Wang et al. 2016). This categorization gives meaning to the stops undertaken by the users and allows for inferences about a user's social relationships (Wang et al. 2016). An algorithm was established that measured the similarity of trajectories while also comparing them to RPC "stop points" which allowed researchers to estimate intimate relationships between users. An additional algorithm was developed to distinguish communities within the trajectories. To accomplish this, each individual trajectory was weighted based on the strength of its similarity to others (Wang et al. 2016). From this, it was possible to infer that teaching buildings were mostly comprised of undergraduates with high mobility and that laboratories have a high graduate student population with low mobility (Wang et al. 2016). These methods for determining relationships were validated using 20 sample pairs of users with existing relationships and 20 sample pairs of strangers. The verification method showed an 89% success rate for estimating intimate relationships.

Instead of attempting to estimate user relationships or roles, this work proposes using established role data to increase the accuracy of a wireless occupancy estimation. It has been made clear by existing studies (Prentow et al. 2015, Ruiz-Ruiz et al. 2014, Wang et al. 2016) that a user's spatiotemporal patterns can determine user-roles. It is the goal of this work to do the opposite; utilize role data to aid in occupancy estimation. Like methods used in (Ruiz-Ruiz et al. 2014, Prentow et al. 2015), this will be done by interviewing library staff regarding the current utilization of the library by university students, staff, and guests. For instance, if it is discovered that library staff members primarily operate during eight-hour shifts each weekday and are typically located on the second floor with low mobility, then the parameters used to filter staff occupancy can be tailored based on that information.

# 3. METHODOLOGY

## *3.1. STUDY AREA*

This work was conducted at Texas State University's Alkek Library (Figure 1), comprised of seven stories and over 27,000m$^2$. Alkek was chosen due to its dense wireless coverage and its large and diverse occupancy rates by campus users. Besides being a library and the largest building on campus (in terms of square footage), Alkek is home to a large teaching theatre, maker space, museum, and more. Detailed floorplans can be found in Appendix A. This section will discuss the following methodologies: 1) retrieval of Wi-Fi log data and preprocessing, 2) occupancy estimation models, and 3) hypothesis testing and validation.

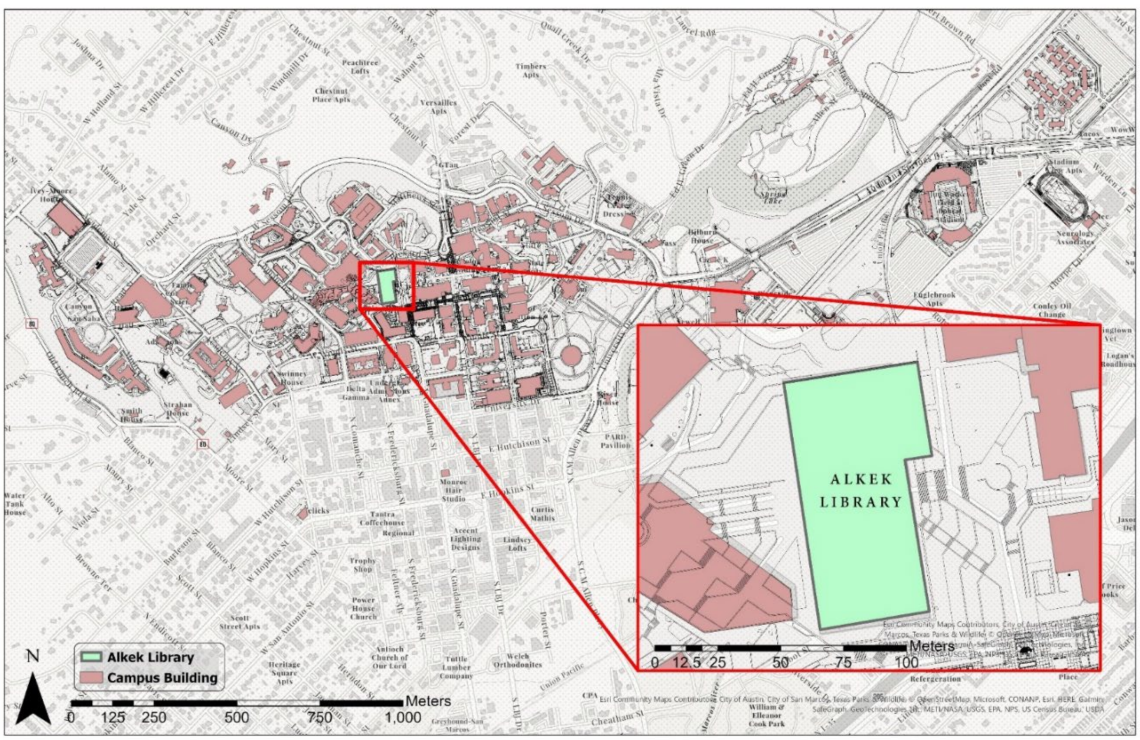



Figure 1. Location map depicting Alkek Library situated at Texas State University.

### *3.2. DATA AND PREPROCESSING*

This work aimed to discover a measure of occupancy based on Wi-Fi log data obtained from Alkek Library. Due to user privacy concerns, a Python script was developed to anonymize usernames and MAC addresses in the log data. The script's principal functionality comes from the Python package Faker, designed to generate synthetic data for privacy-sensitive data. Two weeks of Alkek Wi-Fi log data was collected from a centralized wireless controller and anonymized by Texas State Information Technology (IT) staff before being turned over to researchers for analysis. After receiving the data from the IT department at Texas State University, it was discovered that the Wi-Fi connections associated with users in the Alkek library appear prone to inaccuracies that affect AP connectivity and would substantially influence the various filters proposed to measure occupancy.  These errors can be seen throughout the dataset where a user establishes a seemingly reliable connection to an AP on a given floor, briefly jumps to another floor's AP (e.g., 15-seconds), and then returns to the established floor. These inaccuracies are an inherent limitation of working with Wi-Fi and Bluetooth data and could be due to multipath propagation, signal fading, interference, device roaming behavior, etc. (Elgwad et al. 2019). In the context of our multi-user indoor localization system, floor detection errors can significantly impact the overall accuracy of the location estimation. A robust filter was implemented based on time-based thresholding and one-off values to address this challenge.

#### *3.2.1 DATA SCRUBBING FILTERS*

Two data scrubbing filters were applied to the preprocessed dataset to improve the reliability of Wi-Fi connections. Firstly, a stay-time filter (STF) was created to filter out short, isolated floor transitions, likely caused by erroneous AP connections. The STF filter checks each user's initial floor threshold (IFT) and succeeding floor threshold (SFT). The IFT establishes a user's initial occupancy on a floor given a time threshold. If the IFT is met by exceeding its threshold, and a floor change occurs, the SFT will need to be reached to confirm that a genuine floor change has occurred. If the SFT is not met, the user's floor record will be reverted to the floor indicator established by the IFT. For example, given 1-minute thresholds for both IFT and SFT, let us assume a user connects to the 2$^{nd}$ floor for 4-minutes (satisfying the IFT (> 1-minute)), then connects to the 3$^{rd}$ floor for 45-seconds (not satisfying the SFT (< 1-minute)), and then returns to the second floor. In this case, the SFT was unsatisfied, so all values indicating a change to the third floor would be changed to second-floor values. Alternatively, if SFT is greater than 1-minute in this example, the succeeding floor indicator will be preserved, and the algorithm will continue to iterate through the data. To choose the optimal threshold values for both the initial and succeeding thresholds, the STF was run for every number combination from zero to 3-minutes. Limiting the filters to three minutes was vital to mitigate any offsetting effects of the filter. These effects will be discussed more in the limitations section. For each combination of initial and succeeding floor thresholds, a plot was generated showing the raw floor connections, corrected floor connections, and the manually collected field measurements for "ground truth" as reference (Figure 2).

The reference measurements for this analysis were intermittently collected from Alkek Library over a two-week period. Data were gathered on various floors and sections throughout Alkek Library during the sampling period to mitigate spatiotemporal sampling bias. A researcher visited a floor anywhere from 15-minutes to one-hour while documenting their time and location (current floor). Exact times were recorded using the Timestamp Camera mobile application while entering a new floor through the stairs or elevator. The reference data was added to a table and joined to the researcher's Alkek Wi-Fi logs. Root Mean Square Error (RMSE) was initially adopted to assess the accuracy of various IFT and SFT combinations. However, a potential limitation arises in datasets containing multiple records within short time intervals. The nature of these datasets could lead to inflated RMSE values, rendering the comparisons between two threshold combinations less reliable. This can be seen in the Raw Floor and Corrected Floor column in Table 1 below where the Raw Floor column indicates a floor change to Floor 1 at ID 10 (15:06:32), but the floor change is only realised in the Corrected Floor column at ID 30 (15:09:04). This indicates that a 3-minute offset can result in multiple records not lining up with the raw floor records and could produce misleading results when comparing the effectiveness of the filtering.

Table 1. Subset of STF table output of a single user with IFT of 3-minutes and SFT of three-minutes.

| ID | Datetime | User | AP_Name | Reference Floor | Raw Floor | Corrected Floor |
|---|---|---|---|---|---|---|
| 1 | 2023-03-22T15:05:27.000-0600 | tb1302 | AP4.ALK0.300 | 3 | 3 | 3 |
| 2 | 2023-03-22T15:06:20.000-0600 | tb1302 | AP6.ALK0.300 | 1 | 3 | 3 |
| 3 | 2023-03-22T15:06:20.000-0600 | tb1302 | AP6.ALK0.300 | 1 | 3 | 3 |
| 4 | 2023-03-22T15:06:20.000-0600 | tb1302 | AP6.ALK0.300 | 1 | 3 | 3 |
| 5 | 2023-03-22T15:06:21.000-0600 | tb1302 | AP0.ALK0.579 | 1 | 5 | 3 |
| 6 | 2023-03-22T15:06:21.000-0600 | tb1302 | AP0.ALK0.579 | 1 | 5 | 3 |
| 7 | 2023-03-22T15:06:21.000-0600 | tb1302 | AP0.ALK0.579 | 1 | 5 | 3 |
| 8 | 2023-03-22T15:06:32.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 9 | 2023-03-22T15:06:32.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 10 | 2023-03-22T15:06:40.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 11 | 2023-03-22T15:06:40.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 12 | 2023-03-22T15:06:40.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 13 | 2023-03-22T15:07:00.000-0600 | tb1302 | AP0.ALK0.117 | 1 | 1 | 3 |
| 14 | 2023-03-22T15:07:40.000-0600 | tb1302 | AP0.ALK0.108 | 1 | 1 | 3 |
| 15 | 2023-03-22T15:07:40.000-0600 | tb1302 | AP0.ALK0.108 | 1 | 1 | 3 |
| 16 | 2023-03-22T15:07:41.000-0600 | tb1302 | AP0.ALK0.108 | 1 | 1 | 3 |
| 17 | 2023-03-22T15:07:42.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 18 | 2023-03-22T15:07:43.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 19 | 2023-03-22T15:08:00.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 20 | 2023-03-22T15:08:10.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 21 | 2023-03-22T15:08:10.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 22 | 2023-03-22T15:08:10.000-0600 | tb1302 | AP0.ALK0.145 | 1 | 1 | 3 |
| 23 | 2023-03-22T15:08:20.000-0600 | tb1302 | AP0.ALK0.139 | 1 | 1 | 3 |
| 24 | 2023-03-22T15:08:20.000-0600 | tb1302 | AP0.ALK0.139 | 1 | 1 | 3 |
| 25 | 2023-03-22T15:08:28.000-0600 | tb1302 | AP0.ALK0.133A | 1 | 1 | 3 |
| 26 | 2023-03-22T15:08:28.000-0600 | tb1302 | AP0.ALK0.133A | 1 | 1 | 3 |
| 27 | 2023-03-22T15:08:29.000-0600 | tb1302 | AP0.ALK0.133A | 1 | 1 | 3 |
| 28 | 2023-03-22T15:08:38.000-0600 | tb1302 | AP0.ALK0.133A | 1 | 1 | 3 |
| 29 | 2023-03-22T15:08:38.000-0600 | tb1302 | AP0.ALK0.133A | 1 | 1 | 3 |
| 30 | 2023-03-22T15:09:04.000-0600 | tb1302 | AP0.ALK0.100 | 1 | 1 | 1 |
| 31 | 2023-03-22T15:09:04.000-0600 | tb1302 | AP0.ALK0.100 | 1 | 1 | 1 |

Due to this limitation, a heuristic approach was adopted by visualizing the data and manually inspecting the results of each threshold combination. Plotting the floor transitions and comparing each combination's effectiveness allowed bypassing the pitfalls of relying solely on the RMSE metric. This qualitative approach allowed us to make more informed decisions when determining the appropriate threshold values for our specific dataset, complementing the quantitative evaluation of RMSE. After reviewing each possible combination, it was concluded that an IFT of 1-minute and SFT of 1-minute performed best while mitigating the effects of the potential offset discussed earlier (Figure 2).

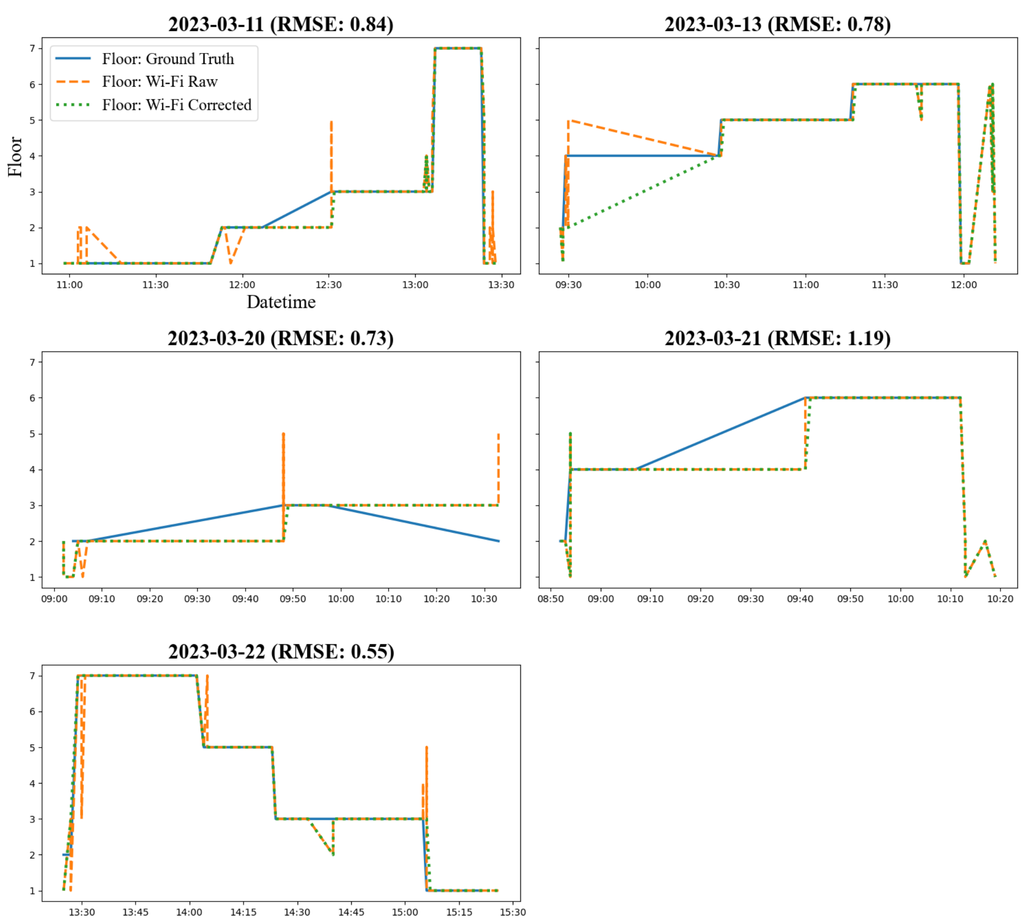



Figure 2. STF with IFT of 1-minute and SFT of 1-minute

Next, a second filter known as the isolated floor filter (IFF) was applied to correct records with one-off floor changes. This works by checking the floor indicator of the previous and following row. If these values are the same and not equal to the current record, the current record is modified to reflect the previous value. For example, consider the following series of numbers: {…2, 2, 3, 2, 2…}. Applying the IFF would change the value three to two as follows: {…2, 2, 2, 2, 2…}. The STF and IFF were applied to each unique user in the dataset.

### *3.3. OCCUPANCY ESTIMATION MODEL*

An initial survey of the APs in Alkek was conducted with the Wi-Fi sniffer application WiFiman. This effort revealed that the initial strategy of accounting for zone-level occupancy estimation using AP propagation would be fruitless due to the high density of wireless APs, physical obstructions, and the sticky nature of the wireless protocol used to connect and disconnect user devices between APs. It was found during the survey that a mobile device would stay connected to a specific AP even while standing under another AP. As mentioned in section 2.2, Anand et al. (2021) experienced similar limitations. Considering the complexity and many mixed-use spaces in Alkek, assigning users to zones based on their role is not feasible in this research. Given this reality, the methods implemented in this work require user devices to be distinguished between floors rather than zones. Through empirical examination of the Wi-Fi sniffer application, it has been determined that moving from one floor to the next provides a reliable break between wireless connections. A device disconnects from an AP when the RSS falls below -75dBm. When this occurs, the device begins roaming, searching for the AP with the highest RSS to connect to.  For example, Figure 3a depicts a mobile device path in green, the expected connections and disconnections in blue, and APs in pink. As a user's device enters Alkek, the mobile phone will begin roaming and connect to AP2.ALK0.200. Although, once the user enters the elevator, the device disconnects. Assume that now that the same user takes the elevator to the fourth floor, the user's device connects to AP5.ALK0.400 upon exiting the elevator (Figure 3b).

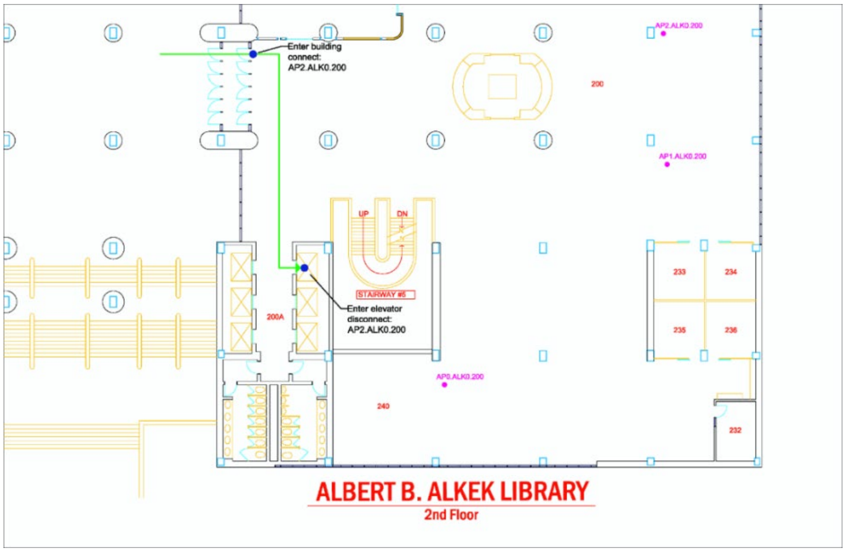



Figure 3a.

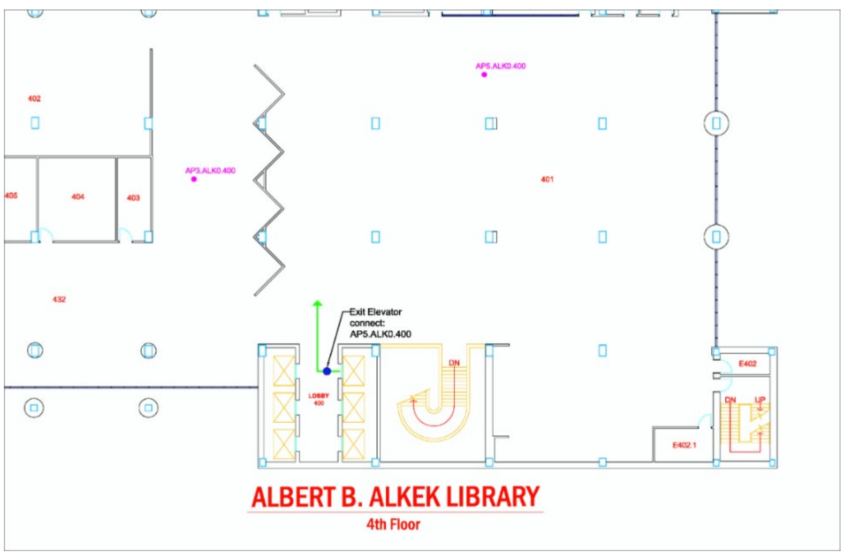



Figure 3b.

Figure 3. Sectional floorplans of Alkek Library on the (a) second floor and (b) fourth floor, where the user's device path is represented as a green line, connection events are blue dots, and APs are pink dots.

The distinction between floors is realized in the data with the AP naming convention, which is broken out into three distinct segments that account for the unique ID of the AP, the abbreviated name of the building the AP is in, and a floor or room number designation (Figure 4). This naming convention is consistent for APs throughout the library.

Figure 4. Wireless access point naming convention


Two methods were used to filter out mobile users. These filters intend to remove users connecting to Alkek APs while passing by the building or moving between floors. As mentioned by Ciftler et al. (2018), differentiating between these brief connections is essential as they could result in overcounting. In conjunction with these two methods, a static device (printers, fax machines, etc.) filter will be applied to remove connections of more than 12-hours that indicate no change in floor-level occupancy (Min et al. 2021). The specific technique used to perform this filter will follow comparable steps to those to be discussed in section 3.3.3.

### *3.3.1 METHOD 1: STATIC MOBILITY FILTER*

The first method employs a uniform -minute filter for all users. This establishes the threshold between mobility and occupancy so that all user devices that maintained a connection with an AP from a distinct floor for 5-minutes or greater will be included in an occupancy count for a given floor. This method of filtering has been documented in several works (Wang et al. 2019b, Wang et al. 2019a, Regodón et al. 2021, Min et al. 2021), with common filtering windows ranging between three (Regodón et al. 2021) and 10-minutes (Wang et al. 2019b). A more detailed explanation of the filtering process is covered below in section 3.3.3. It is also important to note that all methods were be performed during library operating hours to reduce the impact of overdispersion from zero inflation when performing statistical tests (Campbell and O'Hara 2021).

### *3.3.2 METHOD 2: DYNAMIC MOBILITY FILTER*

In method two, additional user-role-specific filters will be applied using expert knowledge of features and services on each floor of Alkek. To retrieve this expert information, a dialog has been established with library staff to obtain a general gauge of a common student, staff, and guest spatiotemporal library usage. Library experts have indicated that most staff members have a standard work schedule of Monday through Friday, 8:00 to 17:00, with only a few staff members working eight-hour shifts (excluding lunch) that either begin at 7:00 in the morning or end at 2:00 at night. This, of course, coincides with the operating hours of Alkek. Additionally, staff presence is significantly reduced on weekends. Library experts also mentioned that their busiest time during the week is at 12:00, corresponding to Figure 5 below.

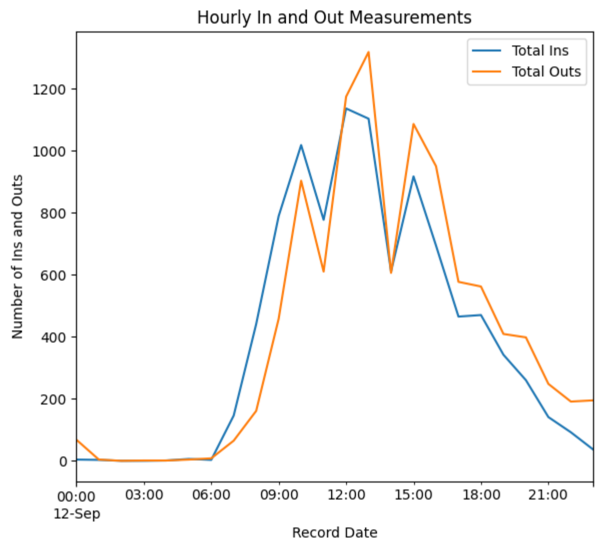



Figure 5. Graph showing Alkek Ins and Outs from RFID data received from Library Staff.

By analysing each floor's themes, features, and services, along with guidance from subject matter experts at Alkek, it is possible to develop a general gauge of expected mobility for students, staff, and guests on each floor. These estimated mobility indicators include Low, Medium (Med), and High, as shown in Table 2. For instance, the first and second floors have the most features and services, indicating a higher mobility rate for students. Although the first two floors are also where most staff offices are located, which would indicate a low mobility rate among staff in these areas. The large number of offices and cubicles in the staff office area insinuates this. Conversely, on every other floor (third-seventh), we expect higher mobility among staff given that they are likely not in their offices but rather performing mobile tasks. Generally speaking, for students, the third, fourth, and seventh floors are used for open study and have a moderate number of features and services. In contrast, the fifth and sixth floors are designated as quiet study floors and have only a few features and no services (Table 2). With these inferences, we can reason that the third, fourth, and seventh floors would have a moderate (or medium) amount of student mobility, while the fifth and sixth floors are expected to have a common measure of student mobility. Additionally, guests' mobility is presumed to be high on the first through sixth floors due to a lack of public features and services.

Table 2. Floor-level breakdown of notable features and services offered at Alkek Library and estimated mobility. Detailed floor plans can be found in Appendix A.

| Floor | Theme | Features | Services | Estimated Mobility |
|---|---|---|---|---|
| 1 | New and Emerging Technology | • DesignSpace<br>• GeoSpace<br>• Immersion Studio<br>• MakerSpace<br>• YouStar Studios | • Alkek Print Shop<br>• ITAC Service Desk<br>• Academic Recording Studios | Student: High<br>Staff: Low<br>Guest: High |
| 2 | Information & Collaboration | • Open Study<br>• Computer Workstations<br>• Teaching Theater<br>• Starbucks<br>• Collaboration Rooms<br>• Library Administration Offices | • Ask Alkek Desk<br>• Checkout, Check-in, Hold pickup | Student: High<br>Staff: Low<br>Guest: High |
| 3 | Academic Research | • Open Study<br>• Public workstations<br>• Scanners | • Teaching & Learning Offices<br>• Research Instruction & Data Services | Student: Med<br>Staff: Low<br>Guest: High |
| 4 | Instruction & Education | • Open Study Areas<br>• Public Lounge<br>• Computer lab<br>• Compact Shelving<br>• Conference Halls<br>• Individual/group study rooms | • Student Learning Assistance Center | Student: Med<br>Staff: Med<br>Guest: High |
| 5 | Quiet & Collaborative Study | • Quiet Open Study<br>• Individual/Group Study Rooms | | Student: Low<br>Staff: High<br>Guest: High |
| 6 | Quiet & Collaborative Study | • Quiet open study<br>• Individual/group study rooms | | Student: Low<br>Staff: High<br>Guest: High |
| 7 | Unique Collections | • The Wittliff Collections Galleries<br>• Open Study<br>• Individual/Group Study Rooms | | Student: Med<br>Staff: Med<br>Guest: Low |

As mentioned above, these occupancy estimates will be used to apply custom parameters for filtering occupancy measurements for each user and floor. Estimated mobility of high, medium, and low will be assigned to a filtering threshold of 5-minutes, 10-minutes, and 15-minutes, respectively. That way, users with high mobilities have shorter filtering window (5-minutes), and those with low mobilities will have longer filtering windows (15-minutes). The researchers assume that having a predefined understanding of library occupancy patterns by various user-roles will aid in calibrating the data parameters and filters that would ultimately improve the accuracy of occupancy estimations.

### *3.3.3 APPLYING FILTERS TO USERS PER FLOOR*

The filters mentioned above are applied to each user based on their duration on each floor. From the Wi-Fi log data (Table 3), floor-level durations will be derived for each user, which will be removed based on the appropriate mobility filter. To distinguish between floors, a regular expression is used to extract all the floor designations from the column "AP Name," shown in Figure 4, and add them to a new column named "Floor" (see Appendix B). Next, in terms of establishing a duration of users per floor, three main problems need to be addressed. Firstly, duplicate records are possible due to multiple network protocols generating the Wi-Fi log data (i.e., Table 3: ID 7 and ID 8). The built-in Pandas function, drop_duplicates(), will be implemented to account for this issue. The second main problem is that data received by IT staff only includes Wi-Fi records associated with the building but not the greater campus at large. This makes it difficult to determine when or if a user has left the building, effectively ending an occupancy session. Thirdly, there can be multiple records for a user on any floor, making it difficult to establish a clear STOP event for each floor when accounting for occupancy duration. For example, three connection logs are associated with user aaa on the first floor in Table 3. In this example, we do not want to represent these three connection logs as three distinct visits, but rather a continuous duration of occupancy.

Table 3. Theoretical snapshot of Wi-Fi log data at Alkek Library (multi-user).

| ID | Datetime | User | Role | AP Name | Floor |
|---|---|---|---|---|---|
| **1** | 2022-01-03 07:21:49-06:00 | aaa | Staff | AP0.ALK.100 | 1 |
| **2** | 2022-01-03 07:21:49-06:00 | bbb | Staff | AP0.ALK.100 | 1 |
| **3** | 2022-01-03 07:22:57-06:00 | ccc | Student | AP0.ALK.100 | 1 |
| **4** | 2022-01-03 07:25:21-06:00 | bbb | Staff | AP1.ALK.102 | 1 |
| **5** | 2022-01-03 07:25:21-06:00 | bbb | Staff | AP1.ALK.102 | 1 |
| **6** | 2022-01-03 07:25:26-06:00 | bbb | Staff | AP0.ALK.100 | 1 |
| **7** | 2022-01-03 07:26:13-06:00 | aaa | Staff | AP0.ALK.100 | 1 |
| **8** | 2022-01-03 07:26:13-06:00 | aaa | Staff | AP5.ALK.108 | 1 |
| **9** | 2022-01-03 07:26:57-06:00 | bbb | Staff | AP5.ALK.108 | 1 |
| **10** | 2022-01-03 07:26:57-06:00 | ccc | Student | AP5.ALK.108 | 1 |
| **11** | 2022-01-03 08:23:22-06:00 | bbb | Staff | AP0.ALK.200 | 2 |
| **12** | 2022-01-03 08:23:22-06:00 | bbb | Staff | AP0.ALK.200 | 2 |
| **13** | 2022-01-03 08:23:44-06:00 | ccc | Student | AP0.ALK.200 | 2 |
| **14** | 2022-01-03 08:23:49-06:00 | ccc | Student | AP0.ALK.200 | 2 |
| **15** | 2022-01-03 08:43:19-06:00 | aaa | Staff | AP2.ALK.200 | 2 |
| **16** | 2022-01-03 08:43:19-06:00 | ccc | Student | AP3.ALK.205 | 2 |
| **17** | 2022-01-03 08:43:51-06:00 | bbb | Staff | AP0.ALK.200 | 2 |
| **18** | 2022-01-03 08:56:27-06:00 | aaa | Staff | AP0.ALK.200 | 2 |
| **19** | 2022-01-03 08:56:27-06:00 | aaa | Staff | AP0.ALK.200 | 2 |
| **20** | 2022-01-03 08:56:29-06:00 | aaa | Staff | AP0.ALK.200 | 2 |
| **21** | 2022-01-03 08:56:31-06:00 | aaa | Staff | AP3.ALK.600 | 6 |
| **22** | 2022-01-03 08:56:32-06:00 | bbb | Staff | AP3.ALK.600 | 6 |
| **23** | 2022-01-03 09:08:16-06:00 | bbb | Staff | AP3.ALK.600 | 6 |
| **24** | 2022-01-03 09:08:39-06:00 | aaa | Staff | AP3.ALK.600 | 6 |
| **25** | 2022-01-03 09:08:52-06:00 | ccc | Student | AP0.ALK.100 | 1 |
| **26** | 2022-01-03 09:08:57-06:00 | bbb | Staff | AP0.ALK.100 | 1 |
| **27** | 2022-01-03 09:10:22-06:00 | aaa | Staff | AP0.ALK.100 | 1 |

To resolve these last two issues, the following methods were applied. The difference between consecutive records was calculated for each user. This time difference was then used to establish a unique visit-ID. Visits to a floor were identified as continuous periods of connectivity lasting at least one-hour. A visit-ID was assigned to each record based on the cumulative sum of the condition that the time difference exceeded one-hour (see Table 4). Fundamentally, the visit-ID accounts for users who have left the building, or disconnected from the network, and the stay-time returns the duration each user spent on each floor. Stay-time is then determined for each user on each floor per visit-ID (See Appendix C). Finally, the non-human and mobility filters are applied for static (see Appendix D) and dynamic methods (see Appendix E).

Table 4. Theoretical creation of the Time Difference and Visit-ID.

| ID | Datetime | User | Role | Floor | Tim | Visit-ID |
|---|---|---|---|---|---|---|
| **0** | 2023-01-01 13:00:00 | aaa | Staff | 1 | | |
| **1** | 2023-01-01 13:15:00 | aaa | Staff | 1 | 00:15:00 | 0 |
| **2** | 2023-01-01 13:30:00 | aaa | Staff | 1 | 00:15:00 | 0 |
| **3** | 2023-01-01 13:45:00 | aaa | Staff | 1 | 00:15:00 | 0 |
| **4** | 2023-01-01 14:00:00 | aaa | Staff | 1 | 00:15:00 | 0 |
| **5** | 2023-01-01 15:30:00 | aaa | Staff | 1 | 01:30:00 | 1 |
| **6** | 2023-01-01 15:45:00 | aaa | Staff | 1 | 00:15:00 | 1 |
| **7** | 2023-01-01 16:00:00 | aaa | Staff | 1 | 00:15:00 | 1 |

Equation (1) will be used to describe the static mobility filter found in section 3.3.1 above, where $C$ is the static occupancy estimation, $f$ is the mobility filtering function described in the above pseudo-code $t_1$ is a universal 12-hour static user threshold and, $t_2$ is a universal 5-minute mobility threshold.

$$C = f(t_1, t_2) \qquad (1)$$

Next, equation (2) describes the dynamic mobility filter found in section 3.3.2 above, where $E$ is the dynamic occupancy estimation, $g$ is the mobility filtering function described in the above pseudo-code, and $t_1$ is a universal 12-hour static user threshold. $t_2$ is a dynamic mobility threshold that is dependent on $r$ and *f*, where r is the user-role and *fl* is the current floor.

$$E = g\big(t_1,\, t_2(r, fl)\big) \qquad (2)$$

*C* and *E* will then be summarized into occupancy estimates to be compared with each other, the field data, and the people-counting sensors, described in the following section, to evaluate the null hypotheses.

### *3.3.4 OCCUPANCY TREND DETECTION*

This work also seeks to utilize the data produced by the RFID-based people-counting sensors found at all three public entrances of the library (see Figure 6).

(a) (b) (c)

Figure 6. Occupancy sensors at the three public entrances of Alkek Library

The data collected from these sensors are aggregated into one-hour intervals representing “Ins” and “Outs” (see Table 5). The two right columns of Table 5 were added to display the ratio of Ins and Outs and estimated occupancy. Estimated occupancy was calculated by subtracting the Total Outs from Total Ins and adding that value to the previously estimated occupancy for each record.

Table 5. Snapshot of occupancy records from all three public entrances combined.

| ID | Location Name | Record Date | Total Ins | Total Outs | Ins/Outs | Estimated Occupancy |
|---|---|---|---|---|---|---|
| **1** | Alkek Library | 2022-09-12 00:00:00 | 4 | 68 | 0.06 | |
| **2** | Alkek Library | 2022-09-12 01:00:00 | 3 | 4 | 0.75 | |
| **3** | Alkek Library | 2022-09-12 02:00:00 | 0 | 0 | 0.00 | |
| **4** | Alkek Library | 2022-09-12 03:00:00 | 0 | 1 | 0.00 | |
| **5** | Alkek Library | 2022-09-12 04:00:00 | 1 | 1 | 1.00 | 0 |
| **6** | Alkek Library | 2022-09-12 05:00:00 | 6 | 4 | 1.50 | 2 |
| **7** | Alkek Library | 2022-09-12 06:00:00 | 3 | 8 | 0.38 | -3 |
| **8** | Alkek Library | 2022-09-12 07:00:00 | 146 | 65 | 2.25 | 78 |
| **9** | Alkek Library | 2022-09-12 08:00:00 | 441 | 161 | 2.74 | 358 |
| **10** | Alkek Library | 2022-09-12 09:00:00 | 789 | 458 | 1.72 | 689 |
| **11** | Alkek Library | 2022-09-12 10:00:00 | 1018 | 903 | 1.13 | 804 |
| **12** | Alkek Library | 2022-09-12 11:00:00 | 777 | 610 | 1.27 | 971 |
| **13** | Alkek Library | 2022-09-12 12:00:00 | 1136 | 1174 | 0.97 | 933 |
| **14** | Alkek Library | 2022-09-12 13:00:00 | 1103 | 1318 | 0.84 | 718 |
| **15** | Alkek Library | 2022-09-12 14:00:00 | 610 | 606 | 1.01 | 722 |
| **16** | Alkek Library | 2022-09-12 15:00:00 | 917 | 1086 | 0.84 | 553 |
| **17** | Alkek Library | 2022-09-12 16:00:00 | 694 | 950 | 0.73 | 297 |
| **18** | Alkek Library | 2022-09-12 17:00:00 | 465 | 577 | 0.81 | 185 |
| **19** | Alkek Library | 2022-09-12 18:00:00 | 470 | 562 | 0.84 | 93 |
| **20** | Alkek Library | 2022-09-12 19:00:00 | 342 | 409 | 0.84 | 26 |
| **21** | Alkek Library | 2022-09-12 20:00:00 | 260 | 398 | 0.65 | -112 |
| **22** | Alkek Library | 2022-09-12 21:00:00 | 141 | 248 | 0.57 | -219 |
| **23** | Alkek Library | 2022-09-12 22:00:00 | 92 | 191 | 0.48 | -318 |
| **24** | Alkek Library | 2022-09-12 23:00:00 | 36 | 195 | 0.18 | -477 |

Library staff provided this data in a daily email to researchers covering the data collection window. Unfortunately, there are two serious limitations of this system. Firstly, there is no way to determine when the library is empty to establish a baseline to begin accounting for occupancy. Secondly, it does not log Ins and Outs at all available entrances/exits. Staff have secure card access to alternative (non-public) entries, commonly used to enter the building at the start of the workday, as they are conveniently located to staff parking. Although it is common for staff to use the public entrances to access other campus buildings, walk to lunch, etc. This creates a discrepancy in the In/Out measurements, making it impossible to measure the occupancy of the building accurately. For example, when calculating occupancy using Table 5, a serious deficit is observed, indicating more Outs than Ins. This can be seen where several of the In/Out values are below one, and the estimated occupancy values fall below zero. Due to these reasons, this method will not be used in place of manually counted reference data to determine the error of wireless occupancy estimation. Instead, the Wilcoxon signed-rank test will compare trends between Alkek's internal RFID people-counter and occupancy estimates generated by Wi-Fi logs. For instance, Figure 5 above shows three distinct peaks at 9:00, 12:00, and 15:00, likely corresponding to typical class schedules.

### *3.4. VALIDATION*

Researchers and volunteers manually collected ground truth reference data to validate the occupancy estimation methods. Each floor had a starting point and was divided in half, with two participants assigned to a given floor at a time. The pair of volunteers would sweep their respective sections and count unique occupants, then retrace their steps generating a second count. The mean of these two counts was calculated for each divided section and added together to create a floor total. This method was repeated for the third – seventh floor from March 2nd, 2023, from 14:01 to 16:35 (Appendix F). reference data were not collected on the first and second floor because they have multiple entrances/exits and restricted areas that researchers cannot access.

To answer the first research question, Hypothesis one is evaluated using the Wilcoxon signed-rank test with a rejection threshold of 0.05 to determine if there is a significant difference between the two filtering methods.

**Hypothesis 1.** Occupancy estimation with dynamic filters derived from user-role data will have a significant difference compared to occupancy estimations without dynamic filters derived from user-role data.

- $\mathbf{H_0}$: There will be **no** significant difference between estimated occupancy with or without dynamic filters.
- $\mathbf{H_a}$: There will be a significant difference between the estimated occupancy with or without dynamic filters.

Next, both occupancy measurements will be compared to manually collected reference data. A Percent Error equation (3) will be employed for each floor measured, where $A_t$ is the actual reference data value and $F_t$ is the forecasted value derived from each model. This will allow for a detailed report of errors for each floor.

$$E = \left|\frac{A_t - F_t}{F_t}\right| \cdot 100 \qquad (3)$$

A more global measurement of error to answer Hypothesis 2 will be calculated using the Mean absolute percentage error (4). Where all variables are consistent with those found in equation (3).

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{At - F_t}{F_t}\right| \cdot 100 \qquad (4)$$

**Hypothesis 2.** Dynamic filters derived from user-role data improve the accuracy of floor-level estimated occupancy when compared to reference data.

- $\mathbf{H_0}$: The mean error % of estimated occupancy with dynamic filters > The mean error % of estimated occupancy without dynamic filters.
- $\mathbf{H_a}$: The mean error % of estimated occupancy with dynamic filters < The mean error % of estimated occupancy without dynamic filters.

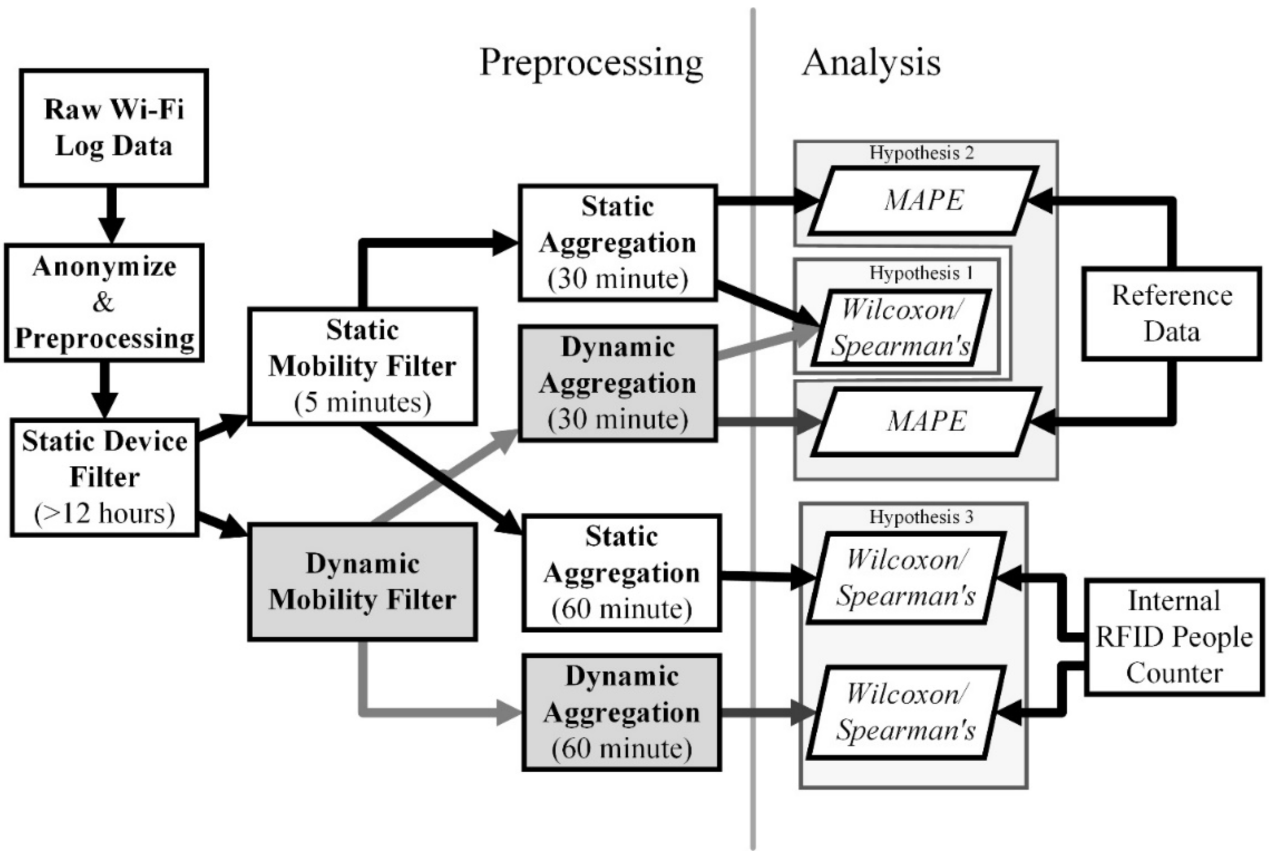
Finally, Hypothesis three will evaluate the significance between the trends of Wi-Fi-derived occupancy estimations and the RFID sensors that currently exist at the public entrances of Alkek Library using the Wilcoxon signed-rank test with a rejection threshold of

0.05. Wi-Fi occupancy estimates will be aggregated into one-hour intervals to match the readings received from the sensor (see Table 5).

> **Hypothesis 3.** Occupancy estimations with and without dynamic filters will have a significant difference when compared to Alkek's internal people-counting sensor.
>
> - **$H_0$**: There will be **no** significant difference between the occupancy estimation methods and sensor data.
> - **$H_a$**: There will be a significant difference between both occupancy estimation methods and sensor data.

The data collection, methodology, and validation above can be visualized in a conceptual model in Figure 7 below.

Figure 7. Conceptual model on methodologies.



## 4. RESULTS

To evaluate Hypothesis one, a Wilcoxon signed-rank test was conducted to compare the two proposed filtering methods (i.e. static vs. dynamic). The resulting counts were aggregated to 30-miniute intervals for comparison. The test yielded a significant difference ($W = 12723.0$; $p < 0.001$) between the static and dynamic occupancy counts. This finding supports the rejection of the null hypothesis, suggesting that occupancy estimation with dynamic filters derived from user-role data differs significantly from estimations without dynamic filters. A graph of total counts for both methods can be seen in Figure 8 below. Furthermore, a strong positive Spearman correlation ($c = 0.979$) was observed between the methods, indicating a consistent relationship. Figure 9 shows a scatter plot highlighting this strong correlation.

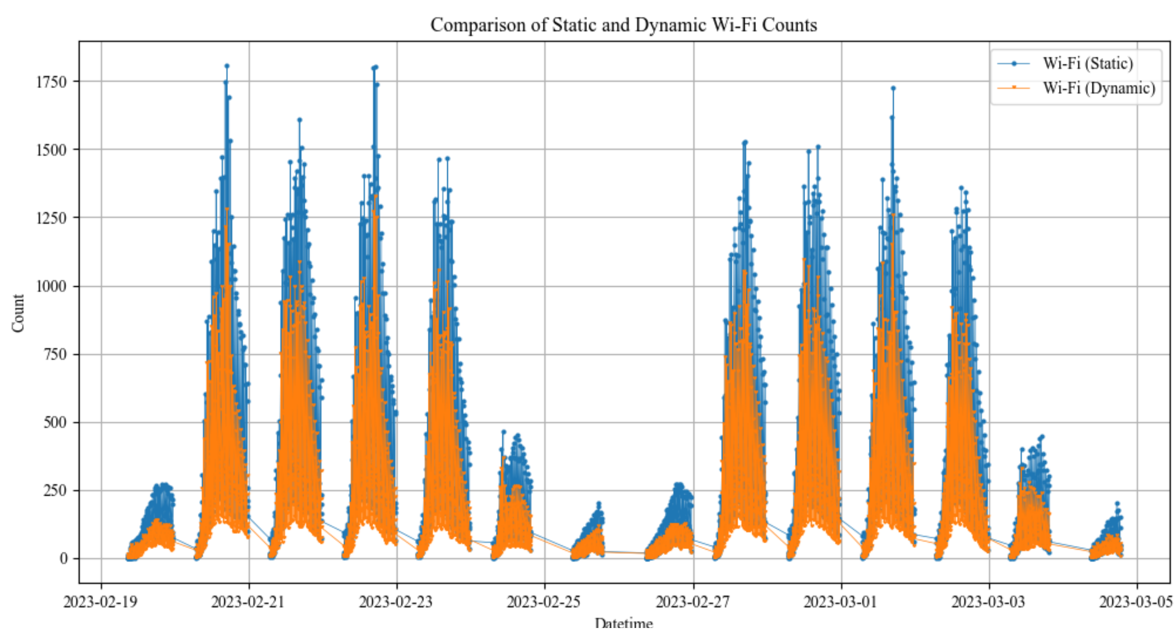



Figure 8. Graph comparing total static and dynamic Wi-Fi Counts (30-minute intervals)

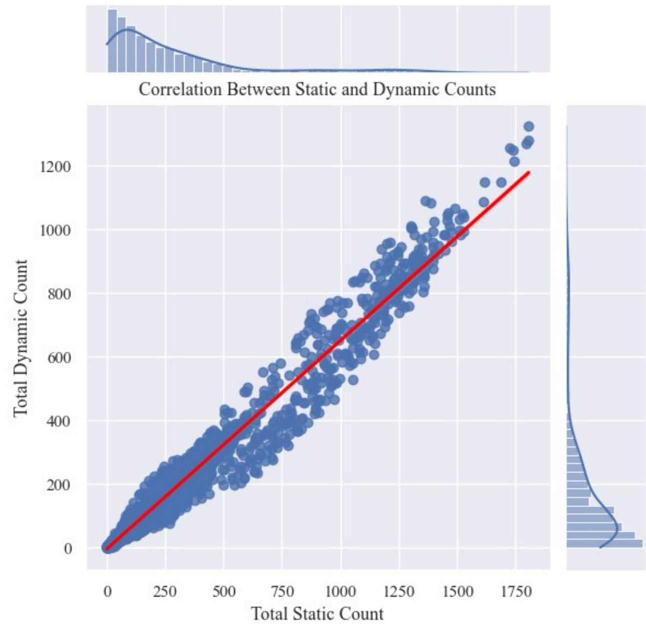



Figure 9. Scatter plot of the correlation between the static and dynamic filtering methods.

Hypothesis two aimed to assess the impact of dynamic filters derived from user-role data on the accuracy of floor-level estimated occupancy compared to reference data. The Mean Absolute Percentage Error (MAPE) was calculated for both the static and dynamic filters aggregated at 30-minute intervals. The static filters yielded a MAPE of 77.28%, while the dynamic filter yielded a lower MAPE of 37.73%. These findings indicate that the dynamic filter outperformed the static filter in terms of accuracy, rejecting the null of Hypothesis two. This is consistent with the floor-level MAPE values for both filter methods in Figure 10 and 11. It is interesting that the Wi-Fi count for both the static and dynamic filters performed significantly worse on the 3$^{rd}$ and 5$^{th}$ floors.

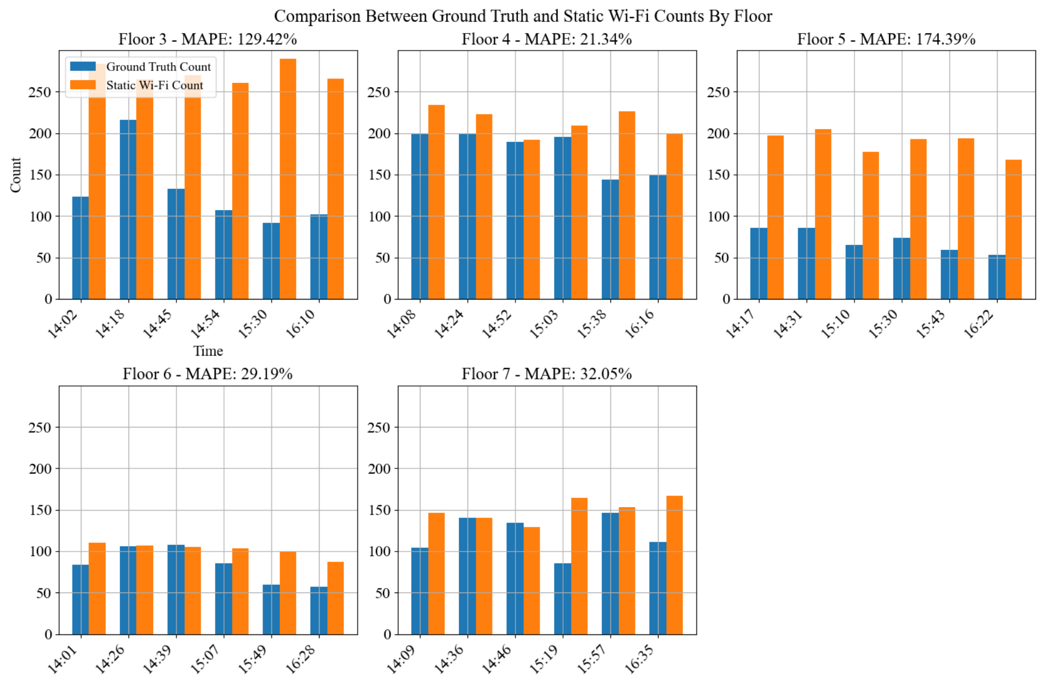



Figure 10. Bar graphs showing the floor-level comparison between reference counts and static Wi-Fi filter counts.

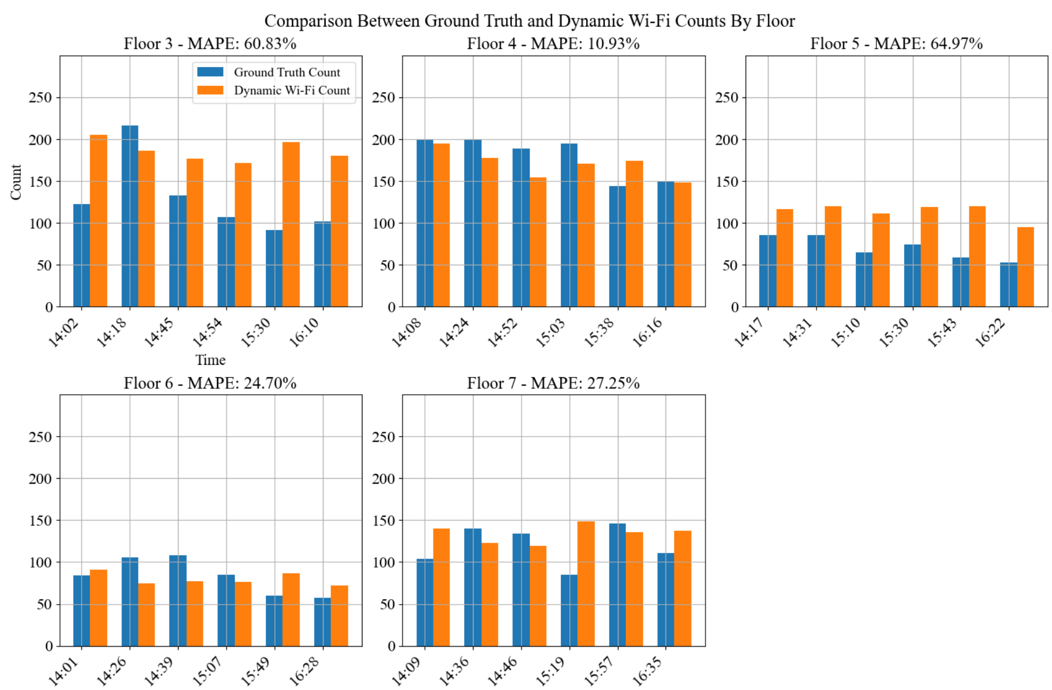
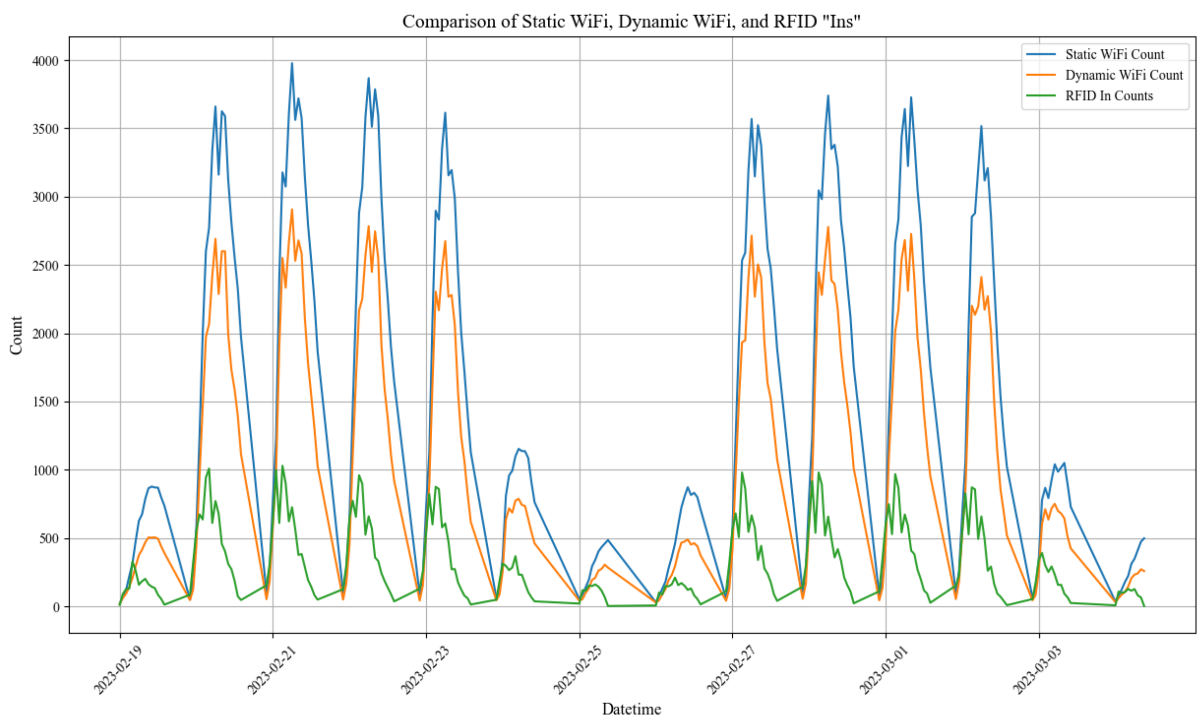



Figure 11. Bar graphs showing the floor-level comparison between reference counts and dynamic Wi-Fi filter counts.

To examine Hypothesis three, the occupancy estimations obtained with the dynamic filters were compared to the Alkek Library's internal RFID people-counting sensor at one-hour intervals. Wilcoxon signed-rank tests were conducted to assess the differences between the RFID sensor and both the static filters ($W = 621.50$; $p < 0.001$) and the dynamic filter ($W = 1365.00$, $p < 0.001$). This indicates a rejection of the null hypothesis, and that both occupancy estimation methods differ significantly from the RFID sensor readings (Figure 12).




Figures 12. RFID people-counting sensor data compared to the counts of the static and dynamic filter (1-hour intervals).

Additionally, Spearman's correlations indicated a moderate positive correlation (correlation: 0.556, $p < 0.001$ between the RFID sensor and the static filter (Figure 13a), as well as between the RFID sensor and the dynamic filter (correlation: 0.619, $p < 0.001$) (Figure 13b).

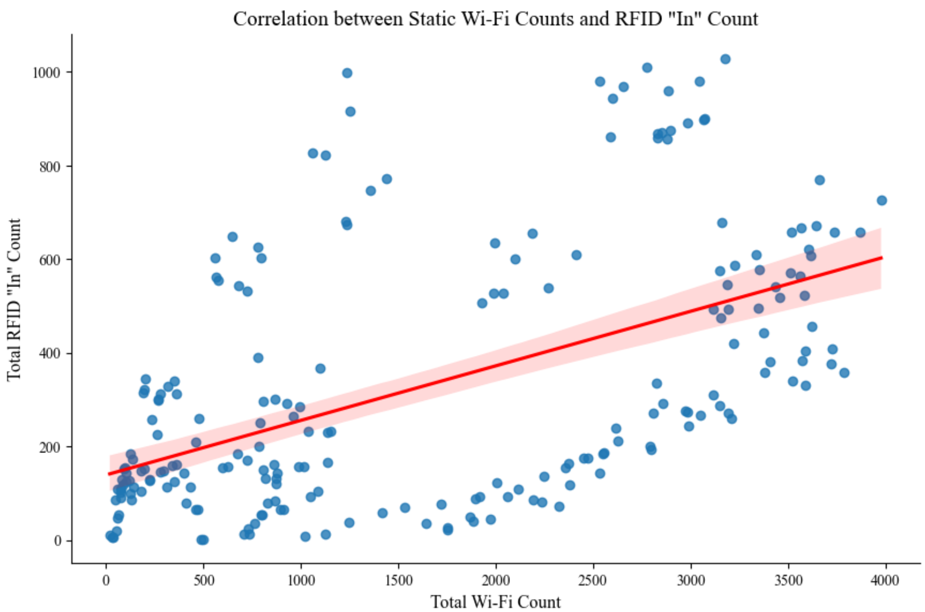



Figure 13a.

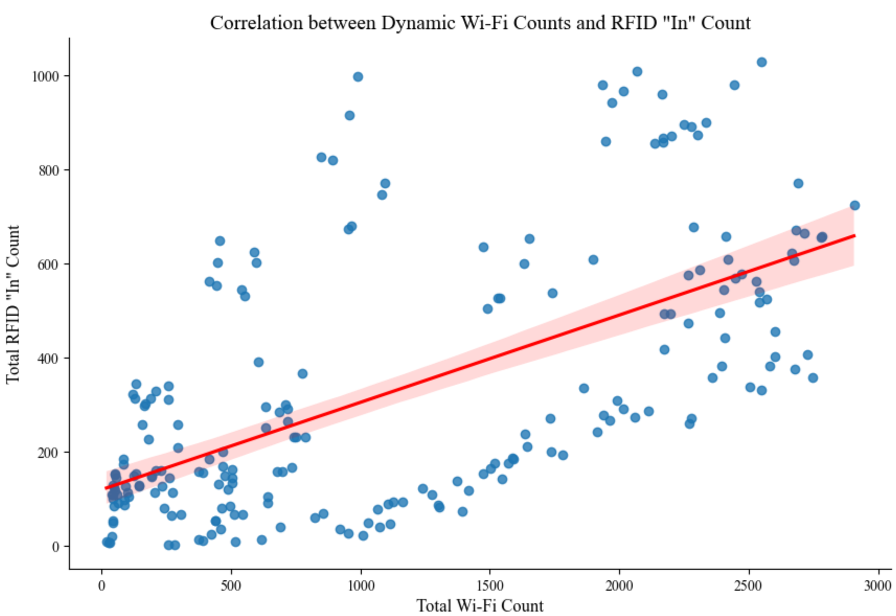



Figure 3b.

Figure 13. Scatter plot showing the correlation between RFID people-counting sensor data and the static (a) and dynamic (b) filter counts.

The rejection of the null hypothesis can be evident to support the postulation that the dynamic filter using the user-role data presents a more effective filter than the static filter for occupancy estimation. The dynamic filter demonstrated improved accuracy compared to the static filter, and both occupancy estimation methods showed significant differences from the RFID sensor readings. These findings underscore the utility of dynamic filters derived from user-role data for estimating floor-level occupancy in a multi-story building and highlight the importance of considering different filtering approaches for accurate occupancy assessments.

## 5. DISUSSION

The results of this work provide valuable insights into the estimations of floor-level occupancy in a multi-story building using a dynamic filter derived from user-role data. In particular, we observed notable differences between the static and dynamic filters in terms of accuracy and alignment with reference data measurements (Figure 10 and 11). However, a deeper analysis is required to understand the factors contributing to these differences and ascertain whether the improved performance of the dynamic filter is solely due to its larger thresholds or other underlying factors. The static filters employed a uniform 5-minute threshold for all users, distinguishing between mobility and occupancy based on a fixed duration. Our findings revealed a constant overestimation of occupancy when using the static filter. This discrepancy may be attributed to the relatively short time threshold, which includes a larger number of transient users who briefly connect to the access points without occupying the floors. The continuous inclusion of the transient users in the occupancy counts likely results in an inflated estimation.

In contrast, the dynamic filter implemented variable thresholds of 5, 10, or 15 minutes, depending on the specific user and floor. This adaptability allowed for a more tailored approach distinguishing between mobility and occupancy based on each floor's services, features, and intended use. Notably, the dynamic filter demonstrated better alignment with the reference data measurements, indicating a more accurate estimation of floor-level occupancy. However, it is essential to acknowledge that the improved performance could potentially be influenced by both the larger thresholds and other factors related to the dynamic nature of the filter. These factors may include variations in user behavior, the distribution of device types across floors, or the presence of specific activities that influence connectivity patterns. Further investigation is warranted to explore these aspects and better understand the specific mechanisms driving the enhanced accuracy observed with the dynamic filters. Additionally, it is worth noting that the larger thresholds of the dynamic filters may introduce a trade-off between accuracy and capturing shorter, intermittent occupancy events. While the dynamic filter demonstrated improved alignment with reference data measurements (Figures 10, 11 and 12), it may potentially overlook short occupancy periods that are of interest in certain scenarios of specific areas within the building. Consequently, the choice between the static and dynamic filter should be carefully considered, taking into account the specific requirements of the occupancy estimation task and desired level of accuracy.

As seen in figure 8 above, the occupancy rates exhibited a consistent weekly pattern. The highest occupancy rates were observed from Monday through Thursday with a peak for both the static and dynamic filter being approximately 1,750 and 1,300 respectively. This aligns with the larger number of classes and therefore higher student and staff activity during these days of the week. On Friday, there was a significant drop in building occupancy. This drop is expected since fewer classes are held on campus on Fridays. This drop may also be influenced by the anticipation of the weekend. Saturday recorded the lowest occupancy levels, with static counts barely exceeding 200, and dynamic counts topping out at around 125. This is consistent with the reduced campus activity on weekends, as most students and staff are not on campus. Sundays saw a moderate uptick in occupancy compared to Saturdays. Static occupancy estimates were just above 250, while dynamic estimates were slightly over 125. This increase is likely due to students and staff preparing for the upcoming week. Notably, the dynamic filter did not exhibit as significant an increase from Saturday to Sunday as the static filter did. This suggests that the dynamic filter, which considers user roles and floor mobility, may be more sensitive to weekends when staff presence is reduced, and fewer classes are in session. Although, it is important to keep in mind that the occupancy estimates examined are a two-week snapshot in time, and that occupancy rates will likely change based on other influences. For instance, weekend occupancy rates may increase during the lead-up to finals.

The observed higher error rates on the 3rd and 5th floors in both the static and dynamic filters (see Figures 10 and 11) could be influenced by several factors. One possible reason for the increased inaccuracies could be the presence of staff offices or other physical obstructions that hindered the volunteers' ability to obtain an accurate reference data count on these floors. It is possible that the presence of these obstructions prevented the volunteers from accurately documenting reference data count. Additionally, if the Wi-Fi signals on the 3rd and 5th floors had a broader coverage area or experienced less attenuation compared to other floors, it could lead to a higher number of users connecting to the APs on those floors. This increased connectivity could introduce noise or inaccuracies in the occupancy estimation process, as users who are only briefly passing through or located in adjacent areas may be counted as occupants. In this case larger thresholds could be applied to these floors to mitigate the issue. One oddity about the notably high error found on the 3$^{rd}$ and 5$^{th}$ floor is that the floors themself have very different features and services. The 3$^{rd}$ floor has more features and services and lends itself to open study and learning space whereas the 5$^{th}$ floor is focused on quite independent and collaborative study. One would suppose floors with similar purposes and therefore similar user spatiotemporal patterns would experience more comparable measures of error. It’s worth noting that the 3rd floor's multifunctional purpose could indeed introduce variations in occupancy patterns. However, the notably high error rate on the 5th floor prompts the consideration of the possibility of additional factors at play, such as cultural or behavioral aspects.  Overall, the observed discrepancies in accuracy on the 3rd and 5th floors between the reference data and the filtered data emphasize the need for careful consideration of the limitations and constraints of the study area when designing and implementing occupancy estimation methods. Although, considering the performance of other floors it stands to reason that by refining the filtering techniques, we can improve the reliability and precision of floor-level occupancy estimations in diverse building environments.

It is important to note that the presence of the stay-time filter (STF) and isolated floor filter (IFF) during the preprocessing stage of this work should also be considered. While the static and dynamic filters were primarily focused on the duration thresholds for mobility and occupancy, the STF and IFF filters targeted specific issues related to Wi-Fi connectivity and floor transitions. The combination of these filters seemed to empirically provide a more reliable estimation of floor-level occupancy by addressing potential sources of error and noise in the dataset.

Based on the comparison between the occupancy estimations obtained with the static and dynamic filters and the Alkek Library's internal RFID people-counting sensor, interesting insights can be gleaned to explain the observed differences. Both the static filters and the dynamic filter produced significantly higher occupancy estimations compared to the RFID "In" count, as indicated by Figure 12 above. Figure 12 takes a broader view of occupancy by presenting data at 1-hour intervals and includes "in" data from the RFID system. Despite the change in interval duration, the occupancy trends remain consistent with those observed in Figure 9. The RFID "in" data, however, tends to show peaks earlier in the day compared to the Wi-Fi filters. This discrepancy can be attributed to the cumulative effects of stay-time, which were calculated for the Wi-Fi-based methods. In contrast, the RFID "in" data solely reflects when people pass through public entrances, capturing foot traffic patterns rather than continuous occupancy. Several factors can contribute to the higher occupancy estimations obtained through the Wi-Fi filters compared to the RFID "In" count. Firstly, it is important to note that the RFID sensor is only located at the three public entrances of Alkek Library and does not cover maintenance or staff entrances (of which there are four). Therefore, individuals entering the library through these entrances would not be accounted for in the RFID count but could still connect to the Wi-Fi network, leading to higher estimations. Secondly, guests who do not possess RFID cards would not be detected by the RFID sensor but could still log into the Wi-Fi network, contributing to the higher occupancy estimations. Similarly, students and staff members might forget their RFID cards or choose not to use them, yet they can still connect to the Wi-Fi network, further increasing the estimations. Lastly, by solely counting "Ins" with the RFID sensor, the duration of occupancy is not considered, and the cumulative effect of individuals staying in the library for extended periods is not captured. This could result in lower counts compared to the Wi-Fi filters, which account for duration in their estimation process. Despite the undercounting observed in the RFID sensor compared to the Wi-Fi filters, it is noteworthy that the pattern of occupancy obtained from the RFID sensor shows similarities to the occupancy patterns derived from the Wi-Fi filters (Figure 12). Both exhibit peaks around noon, indicating increased activity during that time, which adds validity to the overall occupancy trends observed. These findings highlight the advantages of using Wi-Fi filters for occupancy estimation due to their ability to capture a broader range of users, including those entering through non-public entrances or without RFID cards. However, the combined analysis of both methods provides a more comprehensive understanding of the occupancy patterns within the Alkek Library.

In conclusion, this study highlights the impact of filter duration on the estimation of floor-level occupancy. The static filter, with its fixed, 5-minute threshold, consistently overestimated occupancy, while the dynamic filter, with its variable thresholds, demonstrated improved alignment with reference data measurements. However, further investigation is needed to elucidate whether the enhanced performance of the dynamic filter is solely attributed to its larger thresholds or if other factors related to the dynamic nature of the filter are contributing. Ultimately, the choice of the filter should be tailored to the specific context and objectives of the occupancy estimation task, considering the trade-offs between accuracy, and capturing short occupancy events.

## *5.1 LIMITATIONS*

Like most of the research involving Wi-Fi, and other RF technologies, there is an inherent assumption that all users have a wireless device with Wi-Fi functionality enabled. Although, the main limitation observed in this work is the lack of multiple APs and RSS measurements available to researchers, which prohibits the use of more accurate indoor wireless positioning systems such as Fingerprinting. This makes zone-level occupancy estimations difficult. This is compounded by the sticky nature of the AP's network protocol, the dense placement of APs, and the various physical obstructions found throughout Alkek. This reality has restricted the research to floor-level resolution.

Reference data was collected for both tuning the STF and for validating the Wi-Fi occupancy estimations, and each has its own set of limitations that need to be understood. Firstly, the STF reference data was collected by a single researcher using a limited set of device types, including a wireless-enabled Android mobile phone and a Dell XPS laptop, during each walk-through. It is essential to note that different types of devices have varying Wi-Fi antennas, which could affect how they interact with the network, potentially introducing biases in the data.

Additionally, device manufacturers may employ diverse roaming techniques and power-saving options, further influencing data quality. While efforts have been made to mitigate the effects of spatiotemporal sampling bias, it is important to recognize that it may be challenging to account for all possible spatiotemporal patterns that each user may undertake during data collection. This variability could impact the representativeness of the collected data. Lastly, a potential limitation lies in the sample size. The occupancy data was intermittently collected by a single researcher over a two-week period, which may not fully capture the diverse range of user behaviours adequately. A larger and more diverse sample size could provide a more comprehensive understanding of the occupancy patterns in the building.

Limitations are also evident during the collection of reference data used to validate the hypotheses. Firstly, despite the volunteers' careful efforts in data collection, it is reasonable to assume that not all areas of interest within each floor were accurately counted due to practical constraints. These constraints include factors such as the presence of large crowds, locked or restricted areas, and areas unknown to the volunteers. Additionally, the process of collecting ground truth reference data can be arduous and time-consuming, which may lead to fatigue and the possibility of miscounting. This challenge is further compounded by the limited number of volunteers assigned to cover large sections of a given floor.

Significant gaps in connection logs may be present due to battery-saving functionalities found in network devices (Wang et al. 2019a) and lack of user mobility. In Table 6, this can be found when moving from row 6 (T15:05:11) to row 7 (T15:40:30), but more importantly, between row 9 (T15:49:21) and row 10 (T16:38:02) where there is a significant difference between the two logs.

Table 6. Snapshot of Wi-Fi log data at Alkek Library (single user).

| ID | Date/Time | User | Role | RSS | Action | AP Name |
|---|---|---|---|---|---|---|
| 1 | 2022-01-03 14:48:52.000-0600 | tb1302 | Staff | -38 | clientRoaming | AP2.ALK0.200 |
| 2 | 2022-01-03 14:48:52.000-0600 | tb1302 | Staff | | clientAuthorization | AP2.ALK0.200 |
| 3 | 2022-01-03 14:48:57.000-0600 | tb1302 | Staff | | clientInfoUpdate | AP2.ALK0.200 |
| 4 | 2022-01-03 15:05:04.000-0600 | tb1302 | Staff | | clientJoin | AP2.ALK0.200 |
| 5 | 2022-01-03 15:05:04.000-0600 | tb1302 | Staff | | clientAuthorization | AP2.ALK0.200 |
| 6 | 2022-01-03 15:05:11.000-0600 | tb1302 | Staff | | clientInfoUpdate | AP2.ALK0.200 |
| 7 | 2022-01-03 15:40:30.000-0600 | tb1302 | Staff | | clientJoin | AP2.ALK0.200 |
| 8 | 2022-01-03 15:40:30.000-0600 | tb1302 | Staff | | clientAuthorization | AP2.ALK0.200 |
| 9 | 2022-01-03 15:49:21.000-0600 | tb1302 | Staff | | clientAuthorization | AP2.ALK0.200 |
| 10 | 2022-01-03 16:38:02.000-0600 | tb1302 | Staff | | clientRoaming | AP5.ALK0.400 |

From the researcher's perspective, the user could have either shut off their Wi-Fi and remained on the first floor for 50 minutes or left the library and returned 50 minutes later. Either way, this is an inherent limitation when working with RF-based data.

Lastly there are methodological limitations that are important to consider. To start with, the creation of the dynamic mobility filter's threshold values relies heavily on expert knowledge from researchers and library staff to estimate mobility. While this provides valuable insights, it is subjective and may not capture the full range of user behaviours accurately. Additionally, the dynamic filter assumes that users' level of mobility is primarily determined by their roles (e.g., student, staff, guest), though it is important to consider that individual behaviours within these categories can still vary widely. Next, although the dynamic filter method attempts to differentiate each roles estimated mobility (e.g., Low, Med, High) based on each floors features and services, uses' behavior on a given floor may be to complex and dynamic to completely classify. For instance, it would not be out of the realm of possibility for a large group of students to convene on the 2$^{nd}$ floor, which is a high mobility floor for the student role (see Table 2), and perform a low mobility task, such as a study group. Lastly the methodology assumes that user-role based mobility remains consistent over-time, when real-world changes in library operations, policies, or user behaviours may challenge this assumption.

## 6. CONCLUSION

This work introduced methods for measuring floor-level occupancy of a large multi-purpose building at Texas State University using existing wireless infrastructure and coarse Wi-Fi log data. The methods produced provide metrics that aid building management decisions by providing insight into the occupancy patterns of library users. To this end, novel approaches to indoor occupancy estimations have been introduced based on the limitations and unique qualities of the available Wi-Fi log data. The estimation of floor-level occupancy in a multi-story building using dynamic filters derived from user-role data and Wi-Fi logs were explored. The results revealed significant difference between static and dynamic filtering methods in terms of accuracy and correlation with reference data measurements. The dynamic filter, with its adaptable thresholds based on user-roles and floor characteristics, outperformed the static filter, demonstrating improved correlation with the reference data and lower MAPE.

The findings of this study provide valuable insights into the impacts of filter duration on occupancy estimation and underscore the importance of considering different filtering approaches for accurate assessment. However, it is essential to acknowledge the limitation of this research, such as potential inaccuracies when collecting reference data and the inability to achieve precise indoor positioning due to the coarse nature of the data. The presence of visitors and the seemingly unpredictable coverage of the RFID sensor also added complexity to the occupancy estimations.

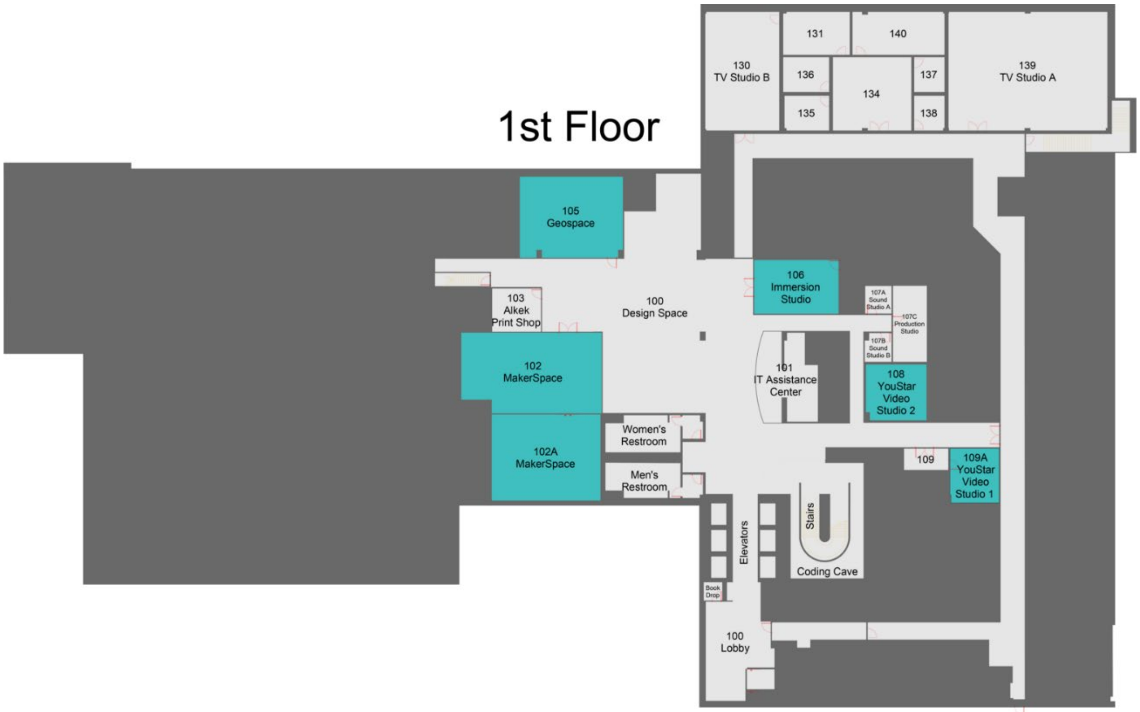
Future work in this area should focus on further understanding the factors contributing to the dynamic filter's improved performance compared to the static filter. Investigating the influence of user behavior, device distribution, and specific activities on the dynamic filter's accuracy. While the dynamic threshold duration used in this study showed promising results at

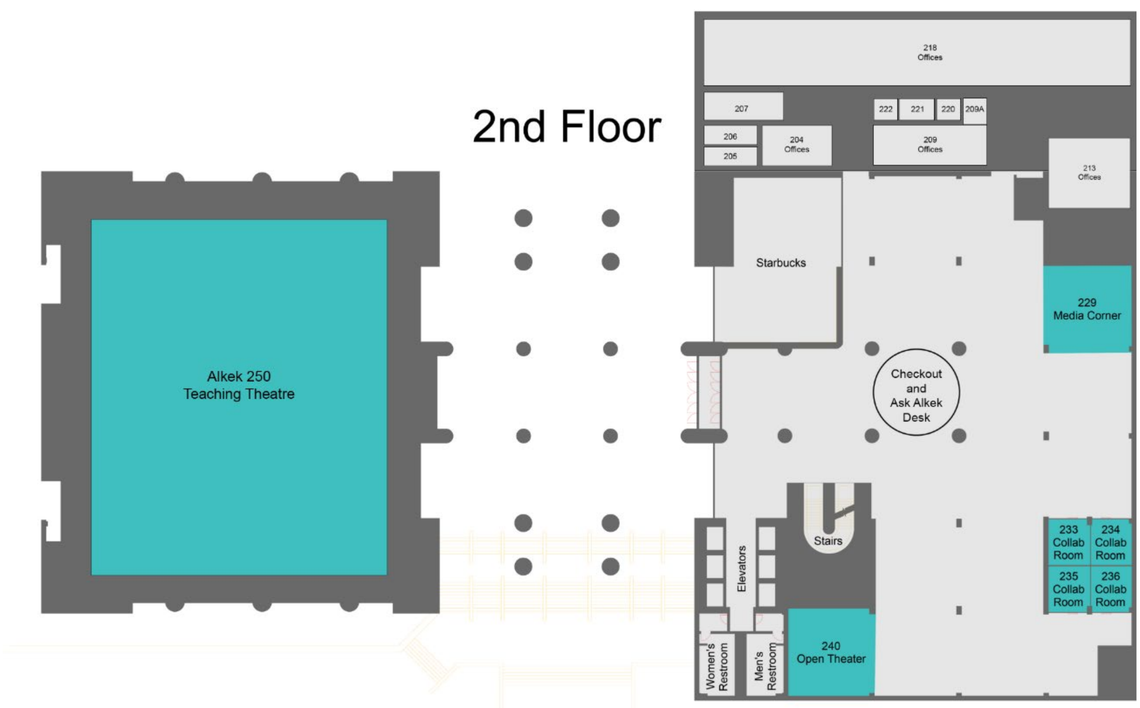
Alkek Library, it is crucial to recognize that its applicability may vary across different buildings. Future research should focus on understanding the factors that influence the optimal threshold durations for accurate occupancy estimations in specific locations and develop methodologies to determine these durations effectively. For example, it may be interesting to apply larger static filters (e.g., 10, 15, 20 minutes) to compare it to their dynamic counterpart. This may introduce additional insight as to how effective the dynamic filter is. Addressing these research questions can enhance our understanding of threshold optimization and contribute to more precise and reliable floor-level occupancy estimations in diverse building environments. Additionally, the observed discrepancies in accuracy on the 3rd and 5th floors between the reference data and the filtered data emphasize the need to investigate these factors comprehensively in future studies. This could result in a more nuanced explanation for the observed variations in accuracy and improve occupancy estimation methods to better suit the unique characteristics of each floor. Furthermore, there is potential to explore the development of real-time indoor occupancy estimation tools using the methodologies introduced in this study. Implementing such tools could assist building management in making informed decisions and optimizing resource allocation based on real-time occupancy patterns. Moreover, advancements in technology and the availability of more comprehensive data sets may enable the application of machine learning algorithms for more accurate and sophisticated occupancy estimations. Moreover, incorporating machine learning models into the filter design could enhance the precision and adaptability of the estimations, taking into account for a wider range of factors influencing user mobility and occupancy.

In conclusion, this study has laid a groundwork for developing an effective occupancy estimation methodology using coarse Wi-Fi log data and user-role information. By addressing the limitations and challenges inherent in Wi-Fi-based estimations, this research contributes to the growing body of knowledge in the field of indoor occupancy assessment and offers valuable insights for building management. Continued research and exploration of innovative techniques will pave the way for more reliable and accurate floor-level occupancy estimations with various resolutions of data and various building environments.

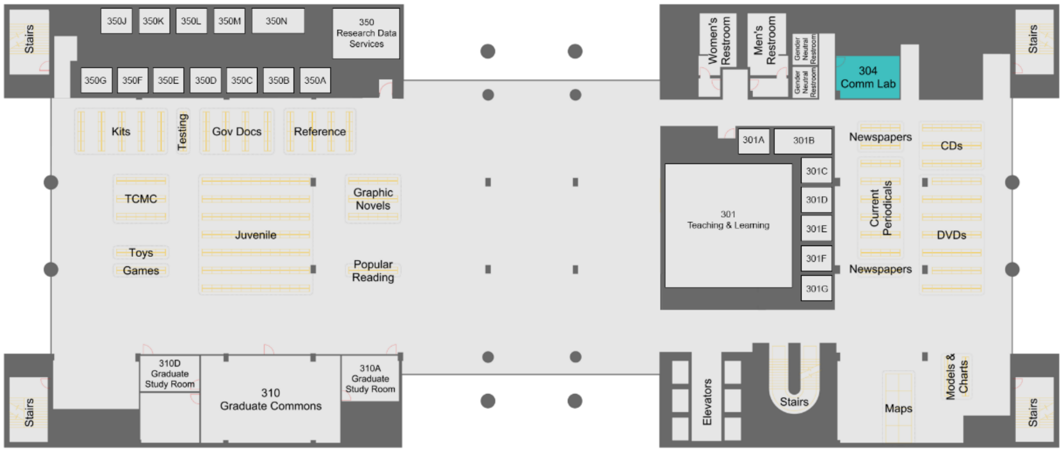
# 7. APPENDIX

A. Detailed Floorplans of Alkek Library

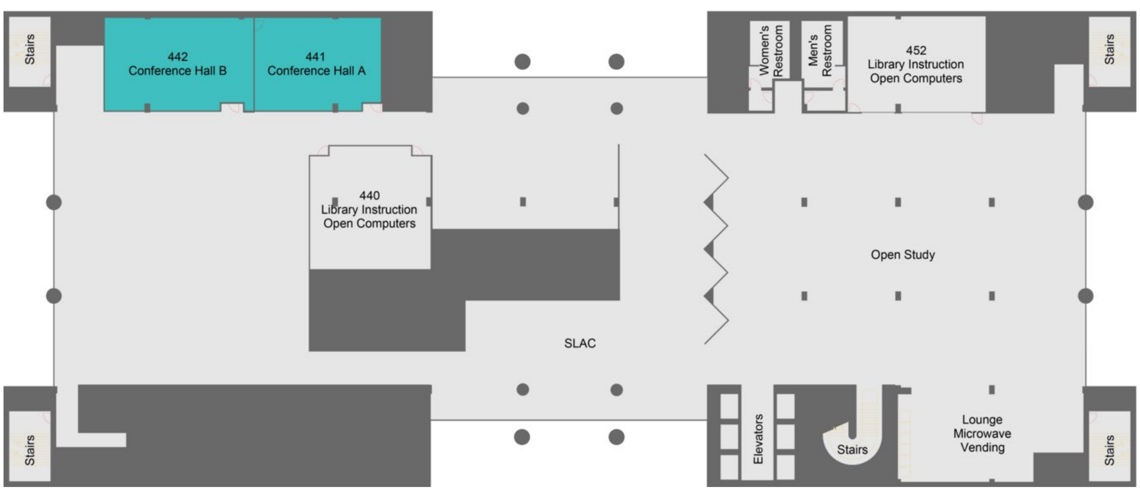


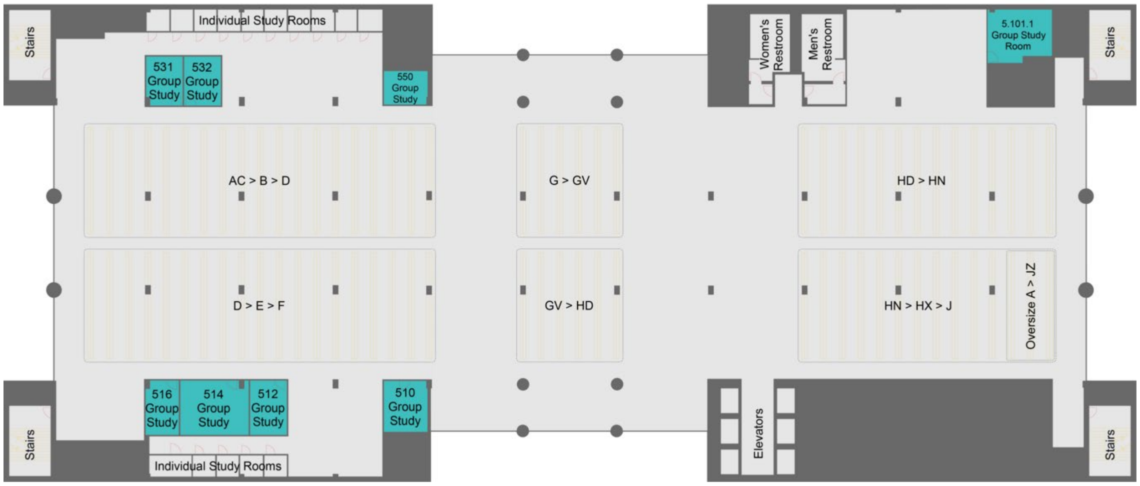
# 3rd Floor



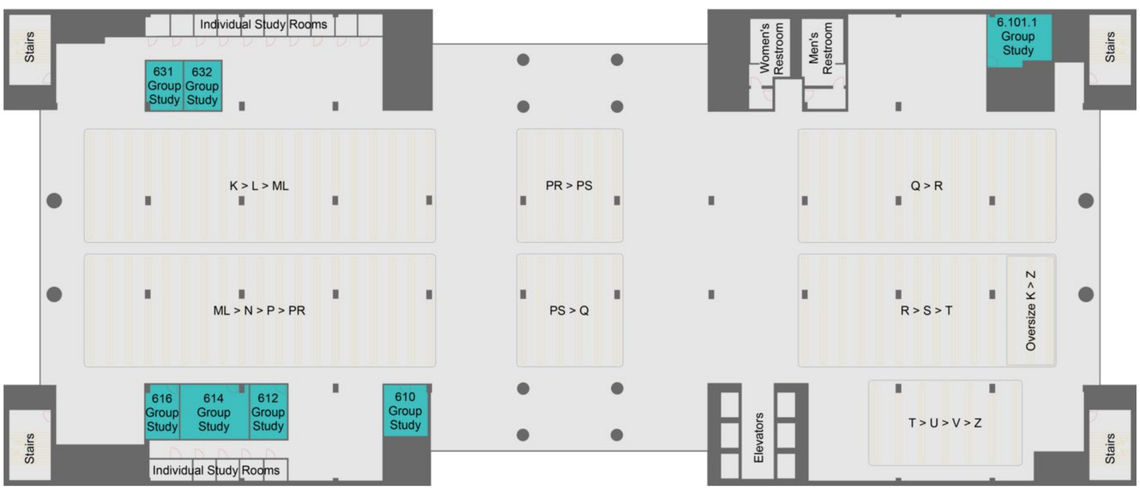

# 4th Floor

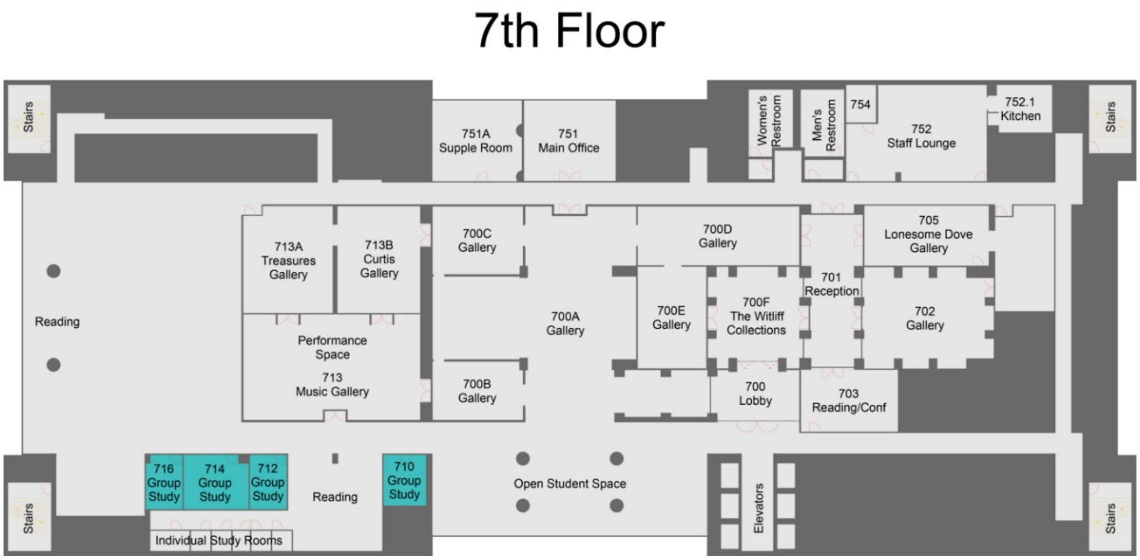


## 5th Floor




## 6th Floor






B. Python script created to assign floor number to each record based on AP name.

```
#select the digit that indicates floor from AP_Name and add it to a
new column "Floor"

df['Floor'] = df['AP_Name'].str.extract('^[^\.]*\.[^\.]*\.[^\.]*?([0-
9])', expand=True)
```

C. Python script detailing the creation of the floor visit-ID and stay-time.

```
# Calculate time difference between consecutive records for each user
df['time_diff'] = df.groupby(['user'])['Datetime'].diff()

# Assign visit-IDs
df['visit_id'] = (df['time_diff'] > pd.Timedelta(hours=1)).cumsum()

# Calculate the stay time for each user on each floor per visit-ID
df['stay_time'] = df.groupby(['user', 'Corrected_Floor',
'visit_id'])['time_diff'].cumsum()
```

D. Python script detailing the Non-Human and static mobility filters.

```
# Non-Human Filter: filter out records with stay times over 12 hours
filtered_data = df[df['stay_time'] <= pd.Timedelta(hours=12)]

# Mobility Filter: filter out records where stay time is less than 5
minutes
filtered_data = filtered_data[filtered_data['stay_time'] >=
pd.Timedelta(minutes=5)]
```

E. Python script detailing the Non-Human and dynamic Mobility filters.

```
# Non-Human Filter: filter out records with stay times over 12 hours
filtered_data = df[df['stay_time'] <= pd.Timedelta(hours=12)]

# mobility lookup table as nested dictionary
mobility_lookup = {
    'Student': {1: 'High', 2: 'High', 3: 'Med', 4: 'Med', 5:
'Low', 6: 'Low', 7: 'Med'},
    'Staff': {1: 'Low', 2: 'Low', 3: 'Low', 4: 'Med', 5: 'High', 6:
'High', 7: 'Med'},
    'Guest': {1: 'High', 2: 'High', 3: 'High', 4: 'High', 5: 'High',
6: 'High', 7: 'Low'}
}

threshold_values = {'High': 5, 'Med': 10, 'Low': 15}

#function to get the mobility threshold based on Floor and vlan_role
def get_mobility_threshold(floor, vlan_role):
    threshold_key = mobility_lookup.get(vlan_role, {}).get(floor,
'High')
    return pd.Timedelta(minutes=threshold_values[threshold_key])

# apply function to store new mobility threshold
filtered_data = filtered_data[filtered_data.apply(lambda row:
row['stay_time'] >= get_mobility_threshold(row['Corrected_Floor'],
row['vlan_role']), axis=1)]

# Group by user, floor, and visit_id, and keep only the first
connection time for each visit
filtered_data = filtered_data.groupby(['user', 'Corrected_Floor',
'visit_id'], as_index=False).agg({
    'stay_time': 'max',
    'Datetime': 'min',
    'vlan_role': 'first'
})
```

F. Table of reference data collected for the third through seventh floors of Alkek Library.

| id | Floor | Datetime | Total |
|---|---|---|---|
| **1** | 3 | 3/2/2023 14:02 | 123 |
| **2** | 6 | 3/2/2023 14:01 | 84 |
| **3** | 4 | 3/2/2023 14:08 | 199 |
| **4** | 7 | 3/2/2023 14:09 | 104 |
| **5** | 5 | 3/2/2023 14:17 | 86 |
| **6** | 3 | 3/2/2023 14:18 | 216 |
| **7** | 4 | 3/2/2023 14:24 | 199 |
| **8** | 6 | 3/2/2023 14:26 | 106 |
| **9** | 5 | 3/2/2023 14:31 | 86 |
| **10** | 7 | 3/2/2023 14:36 | 140 |
| **11** | 6 | 3/2/2023 14:39 | 108 |
| **12** | 3 | 3/2/2023 14:45 | 133 |
| **13** | 7 | 3/2/2023 14:46 | 134 |
| **14** | 4 | 3/2/2023 14:52 | 189 |
| **15** | 3 | 3/2/2023 14:54 | 107 |
| **16** | 5 | 3/2/2023 15:30 | 74 |
| **17** | 4 | 3/2/2023 15:03 | 195 |
| **18** | 6 | 3/2/2023 15:07 | 85 |
| **19** | 5 | 3/2/2023 15:10 | 65 |
| **20** | 7 | 3/2/2023 15:19 | 85 |
| **21** | 3 | 3/2/2023 15:30 | 92 |
| **22** | 4 | 3/2/2023 15:38 | 144 |
| **23** | 5 | 3/2/2023 15:43 | 59 |
| **24** | 6 | 3/2/2023 15:49 | 60 |
| **25** | 7 | 3/2/2023 15:57 | 146 |
| **26** | 3 | 3/2/2023 16:10 | 102 |
| **27** | 4 | 3/2/2023 16:16 | 150 |
| **28** | 5 | 3/2/2023 16:22 | 53 |
| **29** | 6 | 3/2/2023 16:28 | 57 |
| **30** | 7 | 3/2/2023 16:35 | 111 |

## 8. REFERENCES


Anand, P., C. Deb, K. Yan, J. Yang, D. Cheong & C. Sekhar (2021) Occupancy-based energy consumption modelling using machine learning algorithms for institutional buildings. *Energy and Buildings,* 252.

Bi, J., L. Huang, H. Cao, G. Yao, W. Sang, J. Zhen & Y. Liu (2021) Improved Indoor Fingerprinting Localization Method Using Clustering Algorithm and Dynamic Compensation. *ISPRS International Journal of Geo-Information,* 10.

Biczók, G., S. D. Martínez, T. Jelle & J. Krogstie. 2014. Navigating MazeMap: indoor human mobility, spatio-logical ties and future potential. In *IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, 266-271.

Campbell, H. & R. B. O'Hara (2021) The consequences of checking for zero-inflation and overdispersion in the analysis of count data. *Methods in Ecology and Evolution,* 12**,** 665-680.

Chai, W., C. Chen, E. Edwan, J. Zhang & O. Loffeld. 2012. 2D/3D indoor navigation based on multi-sensor assisted pedestrian navigation in Wi-Fi environments. In *Ubiquitous Positioning, Indoor Navigation, and Location Based Service*. IEEE.

Chen, J. & C. Ahn (2014) Assessing occupants' energy load variation through existing wireless network infrastructure in commercial and educational buildings. *Energy and Buildings,* 82**,** 540-549.

Ciftler, B. S., S. Dikmese, I. Guvenc, K. Akkaya & A. Kadri (2018) Occupancy Counting With Burst and Intermittent Signals in Smart Buildings. *IEEE Internet of Things Journal,* 5**,** 724-735.

Danalet, A., B. Michel & B. Farooq. 2012. Estimating Pedestrian Destinations Using Traces from WiFi Infrastructures. In *Pedestrian and Evacuation Dynamics,* eds. U. Weidmann, M. Schreckenberg & U. Kirsch, 1341-1352. Boston, MA: Springer.

Elgwad, A., E. Ashry & B. I. Sheta (2019) Wi-Fi based indoor localization using trilateration and fingerprinting methods. *18th International Conference on Aerospace Sciences & Aviation Technology*.

Fayed, N. S., M. M. Elmogy, A. Atwan & E. El-Daydamony (2022) Efficient Occupancy Detection System Based on Neutrosophic Weighted Sensors Data Fusion. *IEEE Access,* 10**,** 13400-13427.

Jia, M., R. S. Srinivasan & A. A. Raheem (2017) From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews,* 68**,** 525-540.

Khoche, S., K. V. Chandrasekhar, G. V. K. Sasirekha, J. Bapat & D. Das (2021) Occupancy Detection for Emergency Management of Smart Building Based on Indoor Localization. *SN Computer Science,* 2.

Kim, M. & D. Kotz (2005) Modeling users' mobility among WiFi access points. *Workshop on Wireless traffic measurements and modeling***,** 19-24.

Kotz, D. & K. Essien. 2002. Characterizing Usage of a Campus-wide Wireless Network. Hanover, NH, USA 03755: Dartmouth College.

Lee, G., S.-H. Jung & D. Han (2021) An Adaptive Sensor Fusion Framework for Pedestrian Indoor Navigation in Dynamic Environments. *IEEE Transactions on Mobile Computing,* 20**,** 320-336.

Mashuk, M. S., J. Pinchin, P.-O. Siebers & T. Moore (2021) Demonstrating the potential of indoor positioning for monitoring building occupancy through ecologically valid trials. *Journal of Location Based Services,* 15**,** 305-327.

Min, S., R. Ahuja, Y. Liu, A. Zaidi, C. Phu, L. Nocera & C. Shahabi. 2021. CrowdMap: Spatiotemporal Visualization of Anonymous Occupancy Data for Pandemic Response. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, 630-633. Beijing, China.

O'Connor, M. 2021. Discussion regarding nature and availability of Texas State University Wi-Fi log data.

Petrenko, A., A. Sizo, W. Qian, A. D. Knowles, A. Tavassolian, K. Stanley & S. Bell (2014) Exploring Mobility Indoors: an Application of Sensor-based and GIS Systems. *Transactions in GIS,* 18**,** 351-369.

Prentow, T. S., A. J. Ruiz-Ruiz, H. Blunck, A. Stisen & M. B. Kjærgaard (2015) Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring. *Pervasive and Mobile Computing,* 16**,** 305-316.

Rafsanjani, H. N. & A. Ghahramani (2019) Extracting occupants' energy-use patterns from Wi-Fi networks in office buildings. *Journal of Building Engineering,* 26.

Ravi, A. & A. Misra. 2020. Robust, Fine-Grained Occupancy Estimation via Combined Camera & WiFi Indoor Localization. In *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 558-566.

Regodón, A., E. García-Navalón, J. Santiso-Hernandez, E. Delgado-Rodriguez & A. Garcia-Santos (2021) Learnings from User Digital Trail Post-Occupancy Evaluation before COVID-19 for Future Workplace Analysis and Design. *Buildings,* 11.

Ruiz-Ruiz, A. J., H. Blunck, T. S. Prentow, A. Stisen & M. B. Kjaergaard. 2014. Analysis methods for extracting knowledge from large-scale WiFi monitoring to inform building facility planning. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 130-138.

Sapiezynski, P., A. Stopczynski, R. Gatej & S. Lehmann (2015) Tracking Human Mobility Using WiFi Signals. *PLoS One,* 10**,** e0130824.

Schwab, D. & R. Bunt (2004) Characterising the Use of a Campus Wireless Network. *IEEE INFOCOM,* 2**,** 862-870.

Tang, C., W. Li, S. Vishwakarma, K. Chetty, S. Julier & K. Woodbridge. 2020. Occupancy Detection and People Counting Using WiFi Passive Radar. In *IEEE Radar Conference,* ed. IEEE. Washington D.C., USA: IEEE.

Traunmueller, M. W., N. Johnson, A. Malik & C. E. Kontokosta (2018) Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems,* 72**,** 4-12.

Uras, M., R. Cossu & L. Atzori. 2019. PmA: a solution for people mobility monitoring and analysis based on WiFi probes. In *4th International Conference on Smart and Sustainable Technologies (SpliTech)*. University of Split, FESB.

Wang, F., X. Zhu & J. Miao (2016) Semantic trajectories-based social relationships discovery using WiFi monitors. *Personal and Ubiquitous Computing,* 21**,** 85-96.

Wang, J., N. C. F. Tse & J. Y. C. Chan (2019a) Wi-Fi based occupancy detection in a complex indoor space under discontinuous wireless communication: A robust filtering based on event-triggered updating. *Building and Environment,* 151**,** 228-239.
Wang, Y. & L. Shao (2017) Understanding occupancy and user behaviour through Wi-Fi-based indoor positioning. *Building Research & Information,* 46**,** 725-737.
Wang, Z., T. Hong, M. A. Piette & M. Pritoni (2019b) Inferring occupant counts from Wi-Fi data in buildings through machine learning. *Building and Environment,* 158**,** 281-294.
Xia, S., Y. Liu, G. Yuan, M. Zhu & Z. Wang (2017) Indoor Fingerprint Positioning Based on Wi-Fi: An Overview. *ISPRS International Journal of Geo-Information,* 6.