Characterization and analysis of human vertical mobility in complex urban environments: Interplay with horizontal mobility and implications for high-density urban design.

Ajaykumar Manivannan

Thesis submitted to the University of Ottawa in conformity with the requirements for the degree of
Doctor of Philosophy (Ph.D.) in Mechanical Engineering

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering
Department of Mechanical Engineering
University of Ottawa
Ottawa, Ontario, Canada

Abstract

The study of human mobility deals with the movement of human beings (as a group or as an individual) across space and time. Our movements in space affect and are influenced by the built environment, economy, transportation, and the natural environment. These factors are nowhere more intensely interlinked than in cities. Cities are increasingly growing dense and vertical to adapt to the rapid urbanization of our planet. However, a complete understanding of the influence of the vertical dimension on the dynamics of the cities and the lives of its inhabitants needs to be improved. Urban environments are increasingly studied as networks across the three physical dimensions. However, the processes that occur on them, specifically mobility, are still studied as planar (i.e., two-dimensional) processes considering only horizontal mobility. The latter has been extensively studied using various data sources, e.g., census data, travel surveys, call detail records, and GPS. It can therefore be said that there is a general lack of data and methodology to track and derive vertical movements from large-scale human experiments. It is critical for urban planners to understand the patterns of mobility across three physical dimensions and their impact on the built environment. For instance, such knowledge would greatly help improve traffic management, optimize the allocation of facilities and transportation networks, and study several spreading processes (e.g., epidemic diseases).

This research analyzes the patterns of human mobility, encompassing the deeply intertwined horizontal and vertical movements. By studying mobility, valuable insights can be gained, and novel ways can be offered to create, design, and enhance urban built environments with the ultimate goal of prioritizing residents' well-being and quality of life. First, we introduce a city-scale experiment conducted in Singapore, tracking a large population of 50,000 students carrying a wearable device with many sensors. This experiment provides the foundation for accurately deriving vertical displacement based on understanding barometric pressure behavior and developing a ma-

chine learning framework. The second study offers novel contributions to the mobility literature by revealing the spatiotemporal patterns of vertical mobility for the first time, discovering key aggregate statistics of vertical mobility, and showing the intricate interplay and correlation between horizontal and vertical mobility. The third and final study aims to identify and understand the factors that influence user movement in an integrated development, where residential, commercial, and public facilities are distributed in a vertical, high-density built environment. With Kampung Admiralty development as a case study, we show that network-theoretic methods offer a powerful way of looking at data-driven designs at the building scale by associating the network process (user movement) with the network structure (built environment layout).

Preface

The following lists my contributions along with my collaborators for the four results chapters 4, 5, 6, & 7 and the respective publications attached.

- **Chapter 4:**

  Based on publication: Manivannan, A., Chin, W. C. B., Barrat, A., & Bouffanais, R. (2020). On the challenges and potential of using barometric sensors to track human activity. Sensors, 20(23), 6786.

  Contribution statement: A.M. reviewed the literature and wrote the first draft. A.M., W.B., and E.W. conducted the experiment. BB.T. advised on machine learning. A.M, R.B., A.B., and W.B. contributed to the final draft.

  My contributions: I designed and conducted experiments, researched, and wrote the first and final drafts of the manuscript.

- **Chapter 5:**

  Based on publication: Manivannan, A., Willemse, E. J., Chin, W. C. B., Zhou, Y., Tunçer, B., Barrat, A., & Bouffanais, R. (2022). A framework for the identification of human vertical displacement activity based on multi-sensor data. IEEE Sensors Journal, 22(8), 8011-8029.

  Contribution statement: A.M. performed the research and wrote the first draft. R.B., A.B., and E.W. conceived the study. BB.T. advised on machine learning. Y.Z. and B.T. advised on National Science Experiment data. A.M., R.B., A.B., and W.B. contributed to the final draft. All others reviewed the manuscript.

  My contributions: Performed the research and wrote the first and final draft manuscript.

- **Chapter 6:**

Based on manuscript in-preparation: Manivannan, A., Willemse, E. J., Chin, B., Bige, T., Barrat, A., & Bouffanais, R. Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility (under preparation).

Contribution statement: A.M. performed the research and wrote the first draft. R.B. and A.B. supervised the study. A.M., R.B., E.W., and W.B. contributed to the final draft. All others reviewed the manuscript.

My contributions: I performed the research and wrote the first and final draft of the manuscript.

- **Chapter 7:**

  Based on publication: Manivannan, A., Chin, W. C. B., Gopalakrishnan, S., Wong, D. K., Schroepfer, T., & Bouffanais, R. (2024). Effects of the interplay between topology and function of an integrated urban development on patterns of user movement. Scientific Reports, 14(1), 7021.

  Contribution statement: A.M. and W.B. performed the research and wrote the first draft. T.S. and R.B. conceived the experiments. S.G., A.M., W.B., and D.W. conducted the experiments. A.M. and W.B. analyzed the results. W.B., A.M., and R.B. contributed to the final manuscript. All authors reviewed the manuscript.

  My contributions: I designed and conducted experiments, researched, and wrote the first and final drafts of the manuscript.

Acknowledgements

I would like to express my sincere gratitude to all those who contributed to the completion of this research.

I am profoundly grateful to my supervisor, Professor Roland Bouffanais, whose dedication to rigorous scientific inquiry and high standards of excellence has been a constant source of inspiration. His guidance over the course of this thesis has encouraged independent research and cultivated a mindset that aspires toward innovative and thoughtful solutions. His cool and calm demeanor made all the difference.

I would also like to acknowledge the contributions of my collaborators. In particular, I thank Dr. Alain Barrat for his candid observation that "research is hard," a reminder that has consistently motivated critical thinking and persistence. I appreciate the valuable assistance of colleagues and team members—Dr. Srilalitha, Dr. Benny, and Dr. Elias—whose support in empirical study and experimental design was indispensable to this work.

This work was made possible through the financial and logistical support provided by several institutions. I am grateful to CLC and MND for funding the KA pilot project, as well as to Singapore University of Technology and Design (SUTD), Singapore and the University of Ottawa, Canada, for the scholarships that supported my studies. My sincere thanks also go to NSE for granting access to the data crucial for this research.

I want to extend my appreciation to my fellow graduate students and board game guild members (Dr. Sreetaj and Dr. Nikolaj), whose camaraderie and academic discourse helped foster an environment conducive to intellectual and personal growth.

Finally, I wish to acknowledge the support of my wife, Nineesha, whose empowerment and sacrifice have made this timeline possible. My friends, Anand, Abhay, Ashwin, Vijay, Aslem, Rachna, Najla, Ankhita, and Nivey, whose company made me alive.

I dedicate this thesis to my dad, Mr. Manivannan, who is no longer

there to see me finish this, but who planted the seed when I was a child and supported an otherwise unlikely educational and career path through sheer sacrifice and unending love.

Contents

List of Figures

List of Tables

Part I

Introduction and Methods

1

Introduction

1.1 Background and Motivation

The study of human movement has far-reaching applications. Originally focused on migration in socioeconomic contexts, research on human mobility now extends to improving built environment design, telecommunication infrastructure, navigation technologies, transportation facilities, safeguarding public health, and deepening our understanding of human behavior. Measured and modeled human mobility metrics serve as key inputs to these applications; any inaccuracy or incompleteness in our understanding can hinder our ability to design a sustainable future for the planet and its inhabitants.

Rapid advancements in measurement technologies (e.g., high spatiotemporal resolution sensors, wearable devices), modeling techniques (e.g., data analytics, machine learning, artificial intelligence), and scientific disciplines (e.g., complex systems, complex networks, statistical physics) have enhanced our understanding and applications of human mobility. However, significant changes in our built environment have occurred globally over the last few decades, and human mobility studies are still catching up with these trends. Indeed, the study of human mobility is pivotal in understanding and shaping the dynamics of modern urban environments. As cities expand vertically to accommodate growing populations, traditional two-dimensional analyses of

movement—i.e., considering only the built environment parallel to Earth's surface—fail to capture the complexity of human interactions within these spaces. This research addresses this gap by introducing and characterizing movement in the vertical dimension—built environments perpendicular to the Earth's surface—based on several sets of empirical data obtained from human tracking experiments.

Rapid urbanization is a significant global trend, with 55% of the world's population now residing in cities—a figure projected to reach 60% by 2030 [4]. This continuous urban growth has led cities to expand their built-up areas at a rate 50% faster than population growth [5], placing considerable strain on land, energy, and transportation resources. While urbanization presents challenges, studies indicate that innovation, social interactions, and GDP growth increase in tandem with city population; however, crime and disease also trend upward [6]. Interestingly, city infrastructure scales sub-linearly, requiring fewer roads and electric cables per capita [6]. In alignment with the United Nations' Sustainable Development Goals (SDGs), which aim for cities to be inclusive, safe, resilient, and sustainable, a pressing question arises: how can we expand the urban built environment by the projected 1.2 million km^2 by 2030 [5] while adhering to these goals? Urban planners and designers propose a solution: designing for density [7].

Over the past two decades, urban planning has dramatically benefited from extensive studies of human mobility. Gaining insights into human movement is critical for improving traffic management, optimizing land use allocation—including transportation facilities—and enhancing social connectivity. It also aids in understanding the spread of processes such as epidemics [8]. Effectively, how we characterize mobility helps us to model, analyze, understand, and predict the dynamics of a host of key processes. A reciprocal interplay exists between human mobility and the features of the built environment.

High-density cities inevitably expand in the vertical dimension when urban sprawl is constrained by land scarcity or by deliberate design. However, our understanding of how people move vertically is limited. For instance, the average vertical displacement of an urban resident remains unknown. Despite extensive studies on horizontal movements using various data sources (e.g., Call Detail Records (CDRs), Global Positioning System (GPS)), there is a significant gap in understanding vertical mobility. Human mobility has been analyzed almost exclusively as a planar process, which is at odds with the rapid transformation of city topologies—especially in the developing world

and prominently in Asia—where urban environments are increasingly three-dimensional. In most modern cities, high-rise living is the norm. Consequently, human mobility involves numerous transitions between horizontal movement and vertical displacement in residences, transportation facilities, commercial areas, schools, and workplaces. A better understanding of vertical mobility and its interplay with other related features is critically needed to improve quality of life, enhance social interactions, and inform better designs for high-density, livable cities.

1.2 Problem Statement

Lack of large-scale vertical movement data: A fundamental challenge in understanding vertical mobility is the scarcity of large-scale data and methodologies for tracking vertical movement in urban environments. While technologies like GPS enable precise tracking of horizontal movement, there is no equivalent global infrastructure for vertical tracking. Wearable sensors such as micro-electromechanical systems (MEMS) in mobile phones, including barometers, offer an opportunity to address this gap. However, deriving accurate altitude changes from barometric pressure requires substantial data processing, and factors influencing sensor behavior are not fully understood. Existing human activity recognition techniques have advanced in identifying activities like walking and running, but a robust framework for recognizing vertical movements across large populations in urban settings is still lacking.

Unknown characteristics of vertical movement: Large-scale movement data allow us to comprehensively characterize daily commuting patterns in both horizontal and vertical dimensions. Importantly, we also need to study the level of correlation that exists between the two types of mobility. This ensures seamless integration of multi-layer transportation networks, regardless of whether the mode of transport is in the horizontal (e.g., car, train, bus) or vertical dimension (e.g., elevator, escalator, stairs). In addition, statistical characterization of mobility properties, such as the probability distributions of displacement and waiting time, are known to be fundamental mechanisms that drive human mobility. They are also necessary inputs to both human mobility and epidemic disease models. In cities where vertical movements are non-trivial, epidemic or social contagion models often discount contact patterns between humans in elevated spaces. Our inability to model realistic urban movements can thus not only hinder our ability to improve our built environment connections but also impedes our understanding

of human behavior, our efforts to improve population health, and the creation of accurate wayfinding technologies.

Limited quantitative knowledge of factors influencing user movement: There is a lack of comprehensive, empirical understanding of the factors that influence three-dimensional movement in complex urban environments. Elements such as spatial geometry, connectivity, visibility, walkability, spatial function, human familiarity, spatial cognition, location, time, and weather all play a role. Factors like topology—the structure of the built environment's connections—and function—how the space is used—also play critical roles, but quantitative knowledge based on empirical data is sparse and less understood. Such insights are vital for data-driven urban design, especially as cities develop vertically integrated with high-density spaces that combine residential, commercial, and transportation facilities. Vertical transportation facilities can be expected to form the backbone of such developments. Understanding movement patterns in these environments can inform better designs and support initiatives like the development of 15-minute cities/neighborhoods aimed at reducing the carbon footprint associated with mobility.

This thesis thus addresses these three closely related and intertwined problems to enhance our understanding of human movement in cities, with the ultimate aim of improving urban livability and the lives of their inhabitants. These identified gaps present a critical challenge: without understanding vertical mobility, efforts to enhance urban infrastructure, optimize transportation networks, and model spreading processes like epidemics will not have the intended impact.

1.3 Research objectives

This thesis aims to address the identified gaps in understanding vertical human mobility by pursuing the following objectives: (1) to understand and quantify the factors that influence barometric pressure behavior when used to track human activity; (2) to develop a machine learning framework to accurately derive vertical displacement from wearable sensor data; (3) to uncover the spatiotemporal patterns of vertical mobility and its interplay with horizontal movements; and (4) to quantify factors influencing movement within vertically integrated urban developments through a case study. The first two research objectives concern the methodological approach, while the last two relate to empirical contributions. By combining both these approaches, we

aim to provide a comprehensive framework to evaluate vertical movements in high-density cities.

The main research objectives (RO), along with a methodology overview and contributions, are outlined below:

- **RO-1: Understanding and quantifying barometric pressure behavior.** This objective involves conducting a comprehensive literature review of barometric pressure behavior across various scientific fields and performing custom-designed experiments. The aim is to quantify the main factors influencing barometric pressure when a barometer is used to track human activity in diverse urban environments over extended periods. The results will provide crucial insights required to process barometric pressure data to accurately and systematically derive vertical movement. Additionally, this knowledge has far-reaching implications in health monitoring, vehicle tracking, and building monitoring.

- **RO-2: Developing a framework to derive vertical movement.** Building upon the insights from RO-1, the goal here is to design a machine learning framework capable of accurately extracting vertical movement from wearable sensor data, which is often obscured by other daily activities such as being stationary or moving horizontally. Data will be sourced from Singapore's National Science Experiment 2016 (NSE 2016) dataset, which consists of multi-sensor data (e.g., accelerometer, barometer, Wi-Fi localization) recorded by tracking 50,000 students aged 6 to 21 over a week. The resulting framework is expected not only to provide large-scale vertical movement data but also to improve indoor positioning and navigation, estimate energy expenditure, and enhance health monitoring.

- **RO-3: Uncovering vertical mobility and its interplay with horizontal mobility.** This core objective aims to quantify the prevalence of vertical mobility, characterize its complexity through heavy-tailed distributions of key mobility metrics, and examine its interplay with horizontal mobility. The study will utilize the vertical movement data derived from RO-2, combined with a newly derived horizontal movement dataset from the same source. This research will make a foundational contribution to human mobility studies by providing the first large-scale statistical characterization (to the best of our knowledge) of

vertical mobility in the literature. The results highlight the necessity of incorporating vertical mobility into urban planning and transportation models to create sustainable, efficient, and resilient cities that fully embrace their three-dimensional nature.

- **RO-4: Quantifying factors influencing movement within integrated urban developments.** The goal of this case study is to gain a quantitative understanding of key factors—specifically, the topology of the built environment and spatial function—that influence human movement. A data collection experiment has been conducted to track approximately 50 participants over a month as they navigate the Kampung Admiralty (KA) integrated development in the city-state of Singapore. This case study provides an empirical dataset of human movement in mixed-use developments, rarely available with such high spatial and temporal resolution. The effectiveness of using network-theoretic methods to understand movement in three-dimensional environments is assessed. The resulting insights contribute to evidence-based, data-driven design methods aimed at improving the design of high-density, complex urban environments.

  By achieving the said objectives, this thesis intends to serve human mobility researchers by offering novel empirical results, as well as insights for urban planners and designers, transportation researchers, and policy makers.

1.4 Thesis Structure Overview

The thesis is structured as follows: Chapter 2 summarizes the literature reviewed in detail in the appended published and in-preparation articles (see Chapters 4-7); Chapter 3 describes data and methodology in detail, Chapter 4 reviews the factors influencing barometric pressure behavior and the role that barometer sensor plays in tracking human activity, Chapter 5 details the machine framework for deriving vertical displacement; Chapter 6 presents the analysis of vertical mobility patterns; Chapter 7 explores the case study of the Kampung Admiralty integrated development; Chapter 8 provides the overall discussion, implications, and limitations of the thesis results, and concludes with future research directions.

Recommended reading steps for the thesis: The thesis in its entirety can be understood comprehensively by reading Chapter 1 (Introduc-

tion), while Chapters 5 to 7 consist of results in the form of appended articles, and finally ending with Chapter 8, thus skipping Chapters 2 to 4. The literature review summary (Chapter 2) and Chapter 3 Data and Methods are already partially covered in the individual articles. Moreover, Chapter 4 consists of a review article on barometer sensors, which can be skipped for short descriptions relevant to the thesis in Chapter 5.

However, Chapter 2 contains history and background not expanded in the articles but a reader new to human mobility studies may find its content informative. Similarly, Chapter 3 consists of experiments not detailed in the articles, but nevertheless help understand the data acquisition system and improvements of the design of the main data collection experiments, and finally, Chapter 4 comprehensively reviews the barometer sensor that can benefit readers far beyond the scope of the thesis, and factors influencing barometric pressure in detail along with our custom designed experiments.

1.5 Research summary

Each chapter introduces the respective topic in detail with a literature survey, a description of data and methodology, a discussion of results, and finally, summarizes the key results and outlines future works where applicable. The following is a brief outline of the result of chapters 4–7:

- Chapter 4: On the Challenges and Potential of Using Barometric Sensors to Track Human Activity
  - Aim: To review the challenges and potential of using barometric pressure to track human activity, separately and as part of a multi-sensor suite.
  - Challenges: Lack of understanding of barometric pressure behavior when used to track human movements.
  - Novelty: Designed experiments and analyzed the data to understand the factors that influence barometric pressure behavior.
  - Key results: The magnitude and temporal scale of factors influencing barometric pressure are quantified.
  - Publication: The results are published in *MDPI Sensors* [9].
- Chapter 5: A Framework for the Identification of Human Vertical Displacement Activity Based on Multi-Sensor Data.

  - Aim: To derive vertical displacement from a city-scale experiment that tracked a large human population using wearable sensors.
  - Challenges: Lack of methodology and framework to derive vertical displacement from city-scale human tracking experiments.
  - Novelty: Developed a machine learning framework to extract vertical displacement accurately and significantly generalized its applicability to almost all possible real-world conditions.
  - Key results: The developed classifier resulted in 98% overall accuracy and 92% F_1-score in classifying vertical mobility.
  - Publication: The results are published in the *IEEE Sensors Journal* [10].

- Chapter 6: Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility
  - Aim: To characterize and analyze the interplay between horizontal mobility (HM) and vertical mobility (VM).
  - Challenges: (a) lack of city-scale vertical mobility data, and (b) lack of quantitative understanding of even the basic aspects of vertical mobility.
  - Novelty: (a) introduced the mobility data set of $20,000$ subjects based on Chapters 4 and 5, (b) uncovered spatiotemporal patterns of vertical mobility, and (c) showed intricate interplay and correlation between horizontal and vertical mobility.
  - Key results: A subject on the median of the distribution travels 78 meters vertically and 11 km horizontally daily. Vertical mobility is moderately correlated ($\approx 40\%$) with horizontal mobility in the free-range environment (i.e., outside the home and school environments).
  - Publication: Manuscript being finalized.

- Chapter 7: Case study—Effects of the interplay between topology and function of an integrated development on the emergent patterns of user movement
  - Aim: To identify and understand the factors that influence user movement in an integrated development like Kampung Admiralty

(KA) in the city-state of Singapore, where residential, commercial, and public facilities are distributed in a vertical, high-density built environment.

- Challenges: (a) lack of a general framework to analyze the effectiveness of integrated developments, and (b) movement data in integrated developments, or more broadly in mixed-use developments, are sparse.
- Novelty: (a) introduced movement data that tracked 51 participants over a long period of time ($1 \sim 5$ days) in a free-living environment with high temporal (~ 1 Hz) and spatial resolution (~ 20 meters) within a new state-of-the-art integrated development in Singapore, and (b) revealed the quantitative relationship between user movement, spatial and functional components of KA's built environment.
- Key results: Our results revealed a 40% correlation between the centrality measures of the KA spatial network and user movement. Quantified the varying degree of intertwining influence by topology and program on user movement. Vertical streets are found to be the hub of the KA spatial network, forming the main pathway of integrated development.
- Publication: Preliminary works are published as a series of conference presentations and proceedings. The final research article is published in *Scientific Reports* [11]

2

Literature review: A summary

Chapters 4 through 7 comprise individual articles, each with its own self-contained literature review. Therefore, this section serves to provide an overarching summary of the thesis as a whole. Additionally, given the interdisciplinary nature of the research, this section includes detailed background information to assist readers who may be unfamiliar with certain scientific fields or topics that are not extensively covered in the included articles.

In this study, a range of sensors—Bluetooth beacons, accelerometers, Wi-Fi-based localization, and barometers—are utilized to investigate human activity recognition and tracking. While the first three sensors have been extensively researched and are well-understood, the barometer has been explored in a more limited capacity. This thesis aims to enhance the understanding of barometric sensors by combining novel experiments and a thorough literature review to document the factors that influence barometric pressure when tracking human movement. Despite its potential, the barometer has not been fully leveraged in the field of human activity recognition, particularly for measuring altitude changes in urban environments. This research addresses this gap by proposing a framework that integrates barometric and other sensor data to effectively track vertical movement across large populations in city-scale settings. Furthermore, the thesis contributes to the growing body of empirical evidence supporting the concept of volumetric urbanism, which views cities as three-dimensional (physical) spaces, challenging the traditional focus on horizontal urban planning. By examining human mobility, particularly in the vertical dimension, this study highlights its critical role in urban design and public health, advocating for a more comprehensive approach to urban planning. A case study focusing on movement within integrated developments underscores the importance of studying mobility in three physical dimensions to better inform the design of high-density urban environments. The subsequent sections explore the key themes of sensor data behavior, human activity recognition, volumetric urbanism, and human mobility.

2.1 Sensors and sensor data behavior

General Overview of Sensors and Sensor Data Behavior: Sensors play an increasingly critical role in human activity recognition (HAR), which involves using data collected from sensors to identify and track human activities [12]. This field has seen rapid growth with the advent of micro-electromechanical systems (MEMS), which allow the development of small, cost-effective, and energy-efficient sensors like accelerometers, gyroscopes, magnetometers, and barometers. These sensors are commonly embedded in smartphones, wearable devices, and even environmental monitoring systems. Their wide deployment has enabled the collection of massive datasets, allowing researchers and developers to design machine learning models capable of recognizing different activity patterns based on sensor data.

In the context of HAR, the primary goal is to accurately detect and classify activities such as walking, running, sitting, or climbing stairs [12]. This is done by processing and analyzing data from sensors that capture variations in motion, pressure, or environmental factors [13]. For instance, accelerometers measure changes in velocity, which helps in identifying movement patterns. Gyroscopes track rotational movements, while magnetometers help in orientation and direction sensing. These data points provide a comprehensive understanding of human movement and are essential for developing applications that can monitor health, enhance fitness tracking, or optimize urban mobility solutions.

Understanding sensor data behavior is crucial for improving HAR's performance. Raw data collected from sensors often contain noise, variability, and external influences that can hinder accurate classification. Thus, signal processing techniques, such as filtering [14] and data smoothing [15], are necessary to extract meaningful patterns from noisy signals. Additionally, the orientation and placement of sensors play a role in the accuracy of data collection. For example, sensors worn on different parts of the body may capture varying data due to changes in movement dynamics [16, 17]. To address these issues, sophisticated algorithms and classifiers are employed to analyze the data and differentiate between specific activities [18].

HAR systems also benefit from sensor fusion, where multiple sensors are combined to improve accuracy. For example, integrating accelerometer data with barometric pressure readings can lead to better recognition of vertical displacement activities (VDAs) [18, 19], which are typically more challenging to detect with just one type of sensor. This is particularly valuable for applications in urban planning, where tracking vertical mobility (e.g., moving up and down stairs or elevators) helps in understanding how people navigate urban environments. Recognizing these activities is important not only for monitoring daily physical activity but also for applications in public health [20, 21], transportation [22], and infrastructure design.

Barometers and Their Role in Human Activity Recognition: Barometers, traditionally used to measure atmospheric pressure for weather forecasts and as altimeters in aviation, have found new applications in HAR, particularly for tracking vertical movements [15, 23]. MEMS-based barometers are now embedded in smartphones and wearable devices, making them widely available for research and consumer applications. Barometers offer a significant advantage in HAR because they can detect small changes in atmospheric pressure, which correlate directly with changes in altitude. This

makes them particularly useful for recognizing vertical displacement activities (VDA), such as stair climbing, elevator use, and changes in floor levels in buildings.

The role of barometers in HAR has evolved significantly over the past few decades. While accelerometers and gyroscopes are commonly used for detecting horizontal movements and rotational activities, barometers fill a unique gap by providing precise data about altitude changes [18, 24]. This capability is especially important for tracking human activities in complex, multi-level environments like shopping complexes, high-rise residential buildings, or train stations. In these settings, vertical movement is just as crucial as horizontal movement for understanding mobility patterns, making barometers indispensable for a more complete analysis of human mobility behavior.

Barometric pressure data is highly reliable for detecting changes in altitude with an accuracy of about one meter, which is sufficient for distinguishing between different vertical movements [25, 14]. For instance, a change in barometric pressure during an elevator ride can reveal not only the movement but also the speed and direction of the ascent or descent. This can be combined with data from other sensors to improve the overall accuracy of HAR systems. For example, accelerometer data can confirm whether the user is walking or standing still, while the barometer data provides insights into whether the person is moving vertically.

One of the major advantages of using barometers for HAR is their low energy consumption [15]. Unlike accelerometers and gyroscopes, which require frequent data sampling to accurately capture motion, barometers operate efficiently with lower sampling rates, making them more suitable for long-term monitoring applications [26]. This efficiency is critical for wearable devices and smartphones, where battery life is a key concern. Additionally, barometric sensors are less affected by noise and orientation issues, which are common challenges when using accelerometers or gyroscopes alone [23]. This makes barometers an attractive option for applications where accuracy and efficiency are paramount.

Despite their advantages, the use of barometers in HAR is still relatively underexplored compared to other sensors. Research has shown that barometers can significantly improve the accuracy of activity classification, particularly in urban environments where vertical movements are common. However, further research is needed to fully leverage their potential. Integrating barometers with other sensors, such as GPS or Wi-Fi localization, could open up new possibilities for indoor navigation [27] and precise location track-

ing [28], particularly in areas where GPS signals are weak or unavailable, such as underground stations or large buildings.

Factors Influencing Barometric Pressure: While barometers offer valuable data for HAR, several external factors can affect barometric pressure readings, potentially leading to inaccurate activity classification. These factors include changes in climate and weather, the built environment, air velocity during motion, and the inherent accuracy of the barometer sensor itself. Understanding and mitigating these factors is essential for optimizing the use of barometers in HAR.

- **Climate and Weather:** Atmospheric pressure is influenced by weather patterns, temperature, and humidity. Changes in weather, such as storms or high winds, can create fluctuations in barometric pressure that are unrelated to altitude changes. While these variations are typically small (usually less than 5% of absolute atmospheric pressure [29]), they can still introduce noise into the sensor data. For example, a windy day might cause slight pressure differences that could be misinterpreted as vertical movement [30]. However, studies have shown that these effects are generally minor and can be accounted for in data processing algorithms [26, 25]. In particular, the use of differential pressure—calculating the change in pressure relative to a baseline—helps to minimize the impact of weather-related fluctuations [23, 31, 32].

- **Built Environment:** The design and architecture of buildings can also influence barometric pressure readings. For instance, HVAC systems can create pressure differences between indoor and outdoor environments [33]. When a person moves through a door or window, the barometer may register a sudden pressure change that could be mistaken for a vertical movement [34]. This effect is particularly noticeable in highly controlled environments, such as hospitals or clean rooms, where pressure differences are intentionally maintained for safety or cleanliness [21]. Despite these challenges, barometers remain highly effective for tracking vertical movements within buildings, as the pressure changes caused by altitude shifts (e.g., moving between floors) are typically much larger than those caused by HVAC systems.

- **Air Velocity during Motion:** The movement of vehicles, such as buses, trains, or cars, can create pressure fluctuations due to the dynamic interaction between the vehicle and surrounding air. This is

especially pronounced in enclosed environments like tunnels, where the "piston effect" causes rapid changes in pressure as vehicles move through the confined space [35, 36]. These pressure fluctuations can complicate the use of barometers for tracking vertical movement, as the sensor might register these changes as altitude shifts. However, by integrating barometric data with other sensor inputs—such as GPS or accelerometer data—these effects can be filtered out, allowing for more accurate classification of activities.

- **Altitude:** The relationship between altitude and pressure is one of the most predictable factors affecting barometric readings. As altitude increases, atmospheric pressure decreases in a nearly linear fashion, with a drop of approximately 115 Pa for every 10-meter increase in altitude [37]. This predictable relationship allows barometers to accurately track vertical movements, making them highly useful for detecting activities like stair climbing or elevator use. However, variations in altitude can also be influenced by other factors, such as air temperature and humidity, which must be considered when interpreting barometric data.

- **Sensor Accuracy:** The quality of measurement of barometric pressure is limited by the sensor's accuracy. Both absolute and relative barometric pressure can change between devices due to differences in sensors and their characteristics. Absolute pressure can be calibrated to avoid the device dependency [38], while difference in relation pressure measurements between devices can be kept constant by taking into account the noise [39] and drift of sensor data [22]. Other factors to consider include the sensor resolution [31] and sampling frequency, both of which can be customized and chosen based on the spatial and time scale of the activity being recognized.

In summary, while barometers provide a valuable and efficient means of tracking vertical movements, their effective use in HAR requires careful consideration of the external factors that can influence pressure readings. By accounting for these factors and integrating barometric data with other sensor inputs, HAR systems can achieve higher accuracy and reliability, particularly in complex environments like cities or multi-level buildings. This makes barometers a promising tool for future applications in health monitoring, urban planning, and indoor navigation.

In our study, we use the general understanding of sensors and sensor data behavior gained by reviewing the related literature, specifically, about the barometric sensors to design a machine learning framework to derive vertical displacements. A more detailed literature review and our contribution to the documentation of barometric pressure behavior can be found in Section 4 of Chapter 4 (appended article). We also reveal the different spatial and time scales of the different factors that influence barometric pressure, using a mix of literature review and custom designed experiments.

2.2 Human activity recognition

The field of Human Activity Recognition (HAR) heavily relies on sensor data for the identification and classification of specific activities, particularly in complex environments where both vertical and horizontal mobility are significant factors. Over the past few decades, various approaches have been developed to enhance the identification of Vertical Displacement Activity (VDA). This section delves into the current state of research concerning the use of sensors for VDA identification, ranging from simple threshold-based methods to sophisticated machine learning (ML) algorithms.

Accelerometers as the Pioneering Sensor in VDA Studies: One of the earliest applications of sensors in VDA identification was conducted by Sagawa et al. in 1998 [40], who used accelerometers and barometers to identify vertical movement through threshold-based models. The study was limited in scope, involving just six subjects and a small sample time of 83 minutes of activity. Despite these limitations, this work set the stage for future research by demonstrating the potential of sensor fusion for VDA detection. Since then, accelerometers have been the primary sensor used to monitor vertical human activities like stair climbing, escalator rides, and elevator movements. Numerous studies have confirmed the reliability of accelerometers in these tasks [12, 41, 42, 43, 44, 45, 46, 47].

Challenges with Sole Use of Accelerometers: While accelerometers are highly effective in monitoring vertical activities, studies have shown that their classification accuracy is lower compared to horizontal movements [48, 49]. Activities such as stair climbing are particularly challenging to detect with precision, which has led to the integration of additional sensor types to enhance performance. For example, researchers have incorporated gyroscopes, magnetometers, and, more recently, barometers to address these challenges [18, 50, 51, 52, 53]. Liu et al. demonstrated that combining barometer

data with accelerometer and gyroscope data significantly improved classification accuracy from 80% to 90% [18]. This underscores the complementary role that barometers can play when integrated into multi-sensor systems.

Barometers as a Critical Sensor in VDA Detection: Barometers, which measure atmospheric pressure and can detect even slight altitude changes, have emerged as a critical tool in HAR. Their ability to capture vertical displacement makes them indispensable in environments where altitude plays a role, such as multi-story buildings and other vertically built environments. Muralidharan et al. [23] highlighted the robustness of barometers in VDA classification compared to accelerometers, particularly in situations where mobile devices are used for non-activity-related tasks like phone calls. Studies comparing accelerometer-only and barometer-only setups have shown similar classification performances, with barometers being more energy-efficient and less dependent on the user's body position [15].

Challenges in VDA Detection with Barometers: Despite their advantages, barometers are not without limitations. Factors such as air pressure variations, weather conditions, and built environments can affect the accuracy of barometric readings (as described in previous section). Moreover, the classification accuracy of VDA events can vary depending on the resolution of the sensor and the magnitude of vertical displacement. Subtle changes in altitude, for example, may not always be captured accurately by lower-resolution barometers, leading to classification errors. To mitigate these challenges, researchers have developed feature engineering techniques to better extract meaningful data from barometric sensors [54, 18, 14, 26, 55].

Advances in Machine Learning for VDA: The advent of machine learning has brought significant advancements in the identification of human activities, particularly vertical mobility. Traditional threshold-based models have largely been replaced by ML algorithms such as decision trees, random forests, and neural networks [12]. Among these, decision tree models like XGBoost can serve as a highly effective algorithm for VDA classification due to its ensemble learning capabilities. Our study shows that, compared to other algorithms, XGBoost demonstrates superior accuracy in distinguishing between horizontal and vertical displacement activities (HDA and VDA).

In addition to traditional machine learning, researchers have explored the potential of deep learning techniques to improve VDA classification further [56]. Deep learning algorithms, such as convolutional neural networks (CNNs), have been employed to automatically extract features from sensor data, eliminating the need for manual feature engineering [56]. These meth-

ods are particularly useful in complex environments where multiple sensors provide high-dimensional data streams.

Sub-Classifications of Vertical Activities: An emerging area of research within VDA is the sub-classification of vertical activities, such as distinguishing between stair climbing, escalator riding, and elevator movement. To achieve this, a higher sampling rate and more granular data are required, as these activities exhibit different rates of altitude change. Previous studies have shown that sub-classification is feasible with sufficiently high-resolution data [57, 23, 15, 18], although many current datasets, including those used in this study, do not meet these requirements.

In summary, the related works in HAR, particularly concerning VDA detection, reveal a growing interest in leveraging multiple sensors to improve classification accuracy. While accelerometers have traditionally been the go-to sensor for activity recognition, barometers have proven to be invaluable for detecting altitude changes. The integration of machine learning and deep learning techniques has further enhanced the ability to classify human activities in both horizontal and vertical dimensions.

Our study makes an important contribution to HAR (refer to Chapter 5), by developing a framework that recognizes vertical displacement activity with similar accuracy as reported in literature, while significantly expanding the scope of the application (city-wide, large population of $> 16,000$, and tracked over several days). This is done by carefully leveraging the pre-processing methods, developed to identify altitude change in all free-living conditions, by sidestepping other activities or noises with similar signature.

2.3 Study of Human Mobility: History and Applications

This section summarizes the brief history, key applications, and future directions of human mobility studies for readers unfamiliar with the field. Subsequent sections touch directly on this thesis's narrow focus.

Human dynamics is a branch of complex system studies that use tools and techniques of statistical physics to study the movement and interaction of human beings with the overarching aim of understanding human behaviour. The study of human dynamics is remiss without a comprehensive analysis of human mobility, which deals with the movement of human beings (as a group or as an individual) across space and time. Our movements in space affect and are influenced by the built environment, economy, transportation, and the

natural environment. All of these factors co-evolve together as components of an interacting complex system. A vast array of studies is dedicated to studying their parts and gradually understanding their mutual influence on each other.

History of human mobility studies: Human mobility has been studied for centuries, primarily through the lens of migration. In ancient times, climate, weather, and disease have influenced the movement of people from one place to another. As human settlements grew in complexity from distant villages and towns to large cities, people moved for better economic opportunities and adjacency to desired lifestyles. Studies have long documented these travel behaviours and developed theories and models to better understand and predict human mobility and its interaction with other systems of interest.

Prominent theories and supporting empirical studies of human movement emerged in sociology, geography, and economics since the 19^{th} century. Why do we move from one place to another? What influences our choice of destination? What makes a place more attractive than others? Is there a universal pattern behind the human movement? Can we predict human movement on a population or individual scale? After a century of quantitative and qualitative research, these questions still remain elusive. However, many advances- in particular, using new technology- have been made in the last three decades, bringing us closer to answering some of these questions.

With the advent of the digital revolution and advances in microelectronics came the means to track large population movements with high spatial and temporal resolutions. Researchers moved from using empirical data from travel and census surveys and tracking proxy items like dollar bills to studying call detailed records (telecommunication towers) and location data (e.g., GPS, Bluetooth, Wi-Fi). Big data sets on individual mobility attracted physicists and computational scientists to make seminal contributions and advances in the field of human mobility. Today, the study of human mobility is considered an interdisciplinary field that benefits from advances in complex system studies and interweaves with the study of urban complexity [8].

Application of human mobility: The long-term goal of human mobility studies appears to be the development of unifying or universal theories that attempt to predict and explain mobility patterns with all their variability across people, cultures, cities, and countries. Some universal behaviors have been reported in the literature (discussed later in the section) based on global data sets. However, variability in user behavior, built environment,

socio-economic state of humans, and unreliable data sets create challenges to that goal [58]. In spite of this, several applications were found in the pursuit.

The real-world applications of human mobility studies can be categorized under the following topics: urban planning and design, city governance, threat detection to public health, quality of life, communication network, and cybersecurity [8]. Applications in urban planning and design include the optimal development and allocation of resources and facilities in a city or neighborhood and better transportation planning that improves connectivity [59, 60]. Moreover, revealing functional regions or important points of interest also contributes to the better design of cities [61, 62]. An accurate simulation of human movement [63, 64] can improve the prediction of traffic congestion [65, 66] and help with better governance of cities. The most critical application of studying human mobility patterns and modeling its behavior is found in epidemic disease modeling [67, 68, 69]. Studying contact patterns or co-presence networks through the tracking of human movement can better prepare us to detect threats to public health. Besides this, the quality of human life can be improved by air quality monitoring based on mobility traffic [70]. In addition, the social lives of people benefit from an improved recommendation system for travel and other social activities [71, 72]. Studying mobility of urban dwellers along with socio-demographic and socio-economic data can improve our ability to intervene in economic or social isolation of disadvantaged groups and hence increase the quality of life in cities for all. The applications also extend to improvements in engineered systems such as better modeling of the spread of computer viruses with applications to cybersecurity, and improving the communication network or routing protocols through the mobile phones carried by people [63, 73].

Human mobility studies hence cover a vast breadth of research areas that contribute to the above-mentioned applications. The most critical areas of contemporary research can be categorized as follows: (1) understanding and explanation (e.g., factors influencing mobility [74], empirical interpretation or characterization of mobility [75, 76, 77, 78], mobility as a source for understanding other phenomena or systems [62, 79], and revealing the spatiotemporal patterns of mobility [80, 81]), (2) modeling (e.g., modeling human mobility to better explain the underlying mechanisms of empirical observation [82], the pursuit of better models that are in agreement with empirical data [83], and the reproducibility of mobility patterns [84, 80, 85, 86]), and (3) prediction (e.g., predictability of individual or population-level mobility [87, 88, 89], and prediction of next destination). It is worth noting

that general mobility studies also extend to the movement of animals and its impact on ecology [90]. However, our study will focus on human mobility.

Universal laws governing human mobility: An important area of research that attempts to escape the variability and locality of the empirical data is the pursuit of universal laws governing human mobility. Some theories on travel behavior point to invariant properties of mobility such as a constant time budget for travel irrespective of other variables [91], or constant energy consumption for travel [92]. However, both selected examples were later contradicted through data or pointed out for exceptions [8]. Alessandretti et al. [93] analyzed multiple data set with large population tracked over a year to reveal that humans have approximately a constant number of familiar location (~ 25) they visit at any given time. Noulas et al. [94] argue that finding universal laws depends on finding the right variable. By analyzing nearly a million individuals using Foursquare location data, they found that the probability of moving from one location to another is inversely proportional to the number of intervening opportunities between them i.e., displacement is dependent on the rank of the locations rather than the physical distance [94]. Developed based on a multi-city global data set, Schläpfer et al. [95]'s universal visitation law states that the number of visitors to any location decreases as the inverse square of the product of their visiting frequency and travel distance.

Regularity is at the heart of human movements and an important ingredient for finding universal characteristics. Accurate modeling of human mobility and the reproducibility of mobility patterns are contingent upon finding universal laws that govern mobility. Along with the search for universal scaling laws for different metrics of human mobility, these studies represent the forefront of mobility research. Some studies also focus on explaining the origins of the discovered universal laws [96] and attribute this to a diverse set of influences such as population density, urban form, or user preferences. It is worth noting that, although the universal behaviors are based on global and diverse data sets, further replication of these studies is required for robust characterization.

Human mobility data: Human mobility can be studied both from data obtained directly by tracking a human carrying a location sensor and indirectly by tracking a proxy (e.g., taxi [97], airplanes, Twitter, communication channels [59], transportation card, travel survey, and census survey [8]). Direct tracking of humans has the advantage of continuous tracking and potentially higher spatial and temporal resolution. However, direct tracking has

serious privacy implications even at low resolution [98]. On the other hand, tracking human movement by proxy provide high ease for tracking large population albeit with reduced resolution. Some of these indirect sources also lack user-specific (individual) tracking limiting to aggregate studies.

2.4 Human mobility in cities

The core focus of the thesis is the study of human mobility in complex urban environment where density and verticality provide unique challenges. The scope and application of human mobility studies listed in the previous sections remains the same. However, specific dynamics comes into play when mobility is studied within urban boundaries. For example, long displacements (e.g., inter-city or cross-country trips) are ignored, high regularity due to commuting can dominate the mobility patterns, the arrangement of urban centers (e.g., monocentric vs polycentric cities) and the peculiarities of transportation system (e.g., car-centric vs predominant public transportation) and its infrastructure sharply affect the dispersion of humans. In addition, in cities like Singapore (more on what we mean by "cities like Singapore" in the next section), where verticality and density dominates, movement in the vertical dimension can be significant. The chapter 6 of this thesis will thus focus on revealing the importance of studying movement in the vertical dimension by exploring its prevalence, complexity, and integration.

In term of modelling and prediction, the geographic scale of the mobility data has important implications. For example, there is growing research focused on urban mobility that exclusively studies movement within and between cities [95, 59, 94, 97]. Many studies have shown that changing the geographic focus to urban areas changes important metrics of mobility characterization. For example, the probability density function (PDF) fit for the empirical distribution of displacement changes from a power-law to a log-normal or exponential distribution when the geographic scale of mobility data changes from global (or country-wide) to city scale [77]. Liang et al. [96] shows empirically that this is due to the exponential decay in the population density of urban areas. This has profound implications for the modeling and predictability of human mobility. Besides, the focus on urban mobility has led to better answers to urban problems. For example, Xu et al. [82] combined human movement data with urban development data to uncover the fundamental principles behind urban growth. Our study will explore the PDF of key mobility metrics for both horizontal and vertical mobility. This

is first used to benchmark our work with the literature, and second to establish the degree of heterogeneity present in vertical mobility patterns, thus necessitating better understanding for accurate modelling and prediction.

Our urban human mobility study in Chaper 6 aims to focus on two main and interconnected questions regarding vertical movement—(a) How essential is the examination of vertical mobility for comprehending human mobility behaviour? (b) what role does vertical transportation layer play in intra-urban mobility? Answers to these questions will help establish the relevance of vertical dimension to future urban mobility studies.

2.4.1 Singapore city study context

This section provides context to what we mean by "cities like Singapore". We describe Singapore's unique and shared characteristics with other cities and their implications for the generalizability of this human mobility study. In addition, previous human mobility studies conducted in Singapore have also been summarized.

Types of Cities: Cities can be classified based on their urban form (e.g., street networks, horizontal or vertical spread, and density), population or economy size, proportion of public and private transportation, number of spatial centres, walkability, or the geographic distribution of population centres and spatial functions (e.g., business districts, industries, and residential zones). Singapore, while one of the few to have the status of an island city-state, shares common characteristics with several cities. In terms of urban forms, such as density, urban sprawl, and verticality, Singapore is similar to cities like Hong Kong, Shenzhen (China), Guangzhou (China), Seoul (South Korea), Kuala Lumpur (Malaysia), Tokyo (Japan), and Manhattan (United States). On the other hand, Singapore contrasts in the same metrics to cities such as Los Angeles (United States), Phoneix (United States), Canberra (Australia), Copenhagen (Denmark), Ottawa (Canada), Munich (Germany), and Nairobi (Kenya). Our study analyzing vertical mobility patterns in Singapore, therefore, is limited to cities that resemble its vertical and densely built urban environment with strong public transportation infrastructure.

Singapore City: Singapore is an island city-state with a population of 5.9 million residing within a maximum length of 50 km from East to West. Its characteristic features include high verticality and density, predominant public transportation, and a high proportion of public housing (78.7% of population [99] in the form of low-to-high rise apartments built by Singapore government's Housing Development Board (HDB)). Since the 1990s,

Singapore has decentralized its urban form to have multiple self-contained centres (i.e., regions with employment, recreation, and residences). This is different to other cities that have only one central business district, where all transportation and hence traffic flows in and out [100]. Cities with this type of urban form are known as polycentric cities [100]. The city has a central business district in the central region that houses most businesses, followed by other centres spread throughout the island (e.g., Tampines, Jurong East), residential zones spanning the ends of the island, with most industrial zones placed on the end corners [100]. Our study uses the population group of $6 \sim 21$ from ~ 85 schools and higher education institutes spread throughout Singapore to analyze human mobility. For context, primary and Secondary schools are usually found within residential zones, while higher education institutes like polytechnics and universities are sparsely located around the city.

Horizontal transportation in Singapore: Singapore is well connected by its world-class public transportation through trains (260 km of rail lines and 180 stations) and buses ($\sim$ 5000 bus stops). The train stations are placed above and below ground and are typically found every $2 \sim 3$ km. Most of these stations are highly integrated into central shopping complexes and surrounded by residential zones. Of the employed residents, 57.6% took public transportation for commuting, while only 24.1% used cars [99]. Public transportation usage among primary, secondary, and pre-university students are 20.1% (43% walked and 22% used cars), 69.5% (14.4% walked and 13% used cars), and $\sim 90\%$, respectively [99].

Vertical transportation in Singapore: Vertical transportation, especially elevators and escalators, is a common sight in Singapore. Singapore government estimates that the city has more than 70,000 passenger lifts and 7,000 escalators installed (`https://www1.bca.gov.sg/regulatory-info/lifts-escalators/lift-and-escalator-safety`). 95% of Singapore's population live in low-to-high-rise apartments, with only 5% living in landed properties [99]. With few exceptions of very old apartment complexes, almost all are fitted with elevators. All train stations have escalators and elevators, most of which are elevated 2 to 3 floors above or below ground. Most shopping complexes have escalators and elevators as the primary modes of vertical transportation, with stairs placed and accessible for emergency exits only. Bridges for pedestrian road crossings are fitted with elevators for elderly accessibility, and it is now increasingly common to have one-passenger-wide escalators. The population studied is likely to access stairs only in low-rise

schools. At the same time, all homes and publicly accessible places like public transportation and shopping centres are more likely to be traversed through motorized vertical transportation. This infrastructure gave rise to our rich vertical transportation data to comprehensively analyze vertical human mobility patterns in Singapore.

Human mobility studies in Singapore: Human mobility (horizontal) has been extensively studied in Singapore. The source of data, methodology, and applications used resemble human mobility studies globally. The sources include transit card data [101, 102], call detailed records [103, 104, 102], GPS, social network data (e.g., Foursquare [94], Twitter [105]), surveys [100], and focus groups [106]. The studies had applications to public health (e.g., epidemic diseases [107, 104], air quality [108]), urban planning and design [104, 100, 105], universal laws in human mobility behaviour [94, 95], transportation planning [101, 102, 106], and humanities [106] (e.g., the influence of socio-economic factors [103]).

Singapore's mobility data has also been used in prominent human mobility studies that helped shape the narrative of the field, especially in discovering universal laws in human mobility. Noulas A et al. [94] compared 34 cities, including Singapore, to show that the distribution of rank-distance is universally common among all countries better than the distribution of displacement. They also showed that higher density in cities (meaning more opportunities nearby) leads to lower mean displacement (distance travelled) among the population. Schläpfer et al. [95] analyzed the Greater Boston area and Singapore to show that a universal visitation law exists where "the number of visitors to any location decreases as the inverse square of the product of their visiting frequency and travel distance". Both these studies reaffirmed long-standing theories in Geography, such as Stouffer's theory of intervening opportunities [94], central place theory and Weber's theory of emergent optimality [95]. This shows that, by picking the right mobility indicators, a comparison of urban mobility behaviour can be made between any two cities.

2.5 Human mobility in integrated development

The study of human movement in urban environments has evolved with the growing interest in integrated developments. Mixed-use developments, where residential, commercial, and public spaces co-exist within a closely-knit, well-planned framework, have emerged as a possible solution to address the chal-

lenges arising from rapid urbanization [7, 109]. These developments optimize land use by increasing density and minimizing the environmental impact of transportation. However, research on mixed-use developments remains sparse [110, 111, 112]. Despite limited studies, there is growing consensus that mixed-use developments could play a key role in promoting urban sustainability. Integrated developments—mixed-use developments closely tied to transportation networks—offer an even broader scope, presenting potential improvements in social interactions, reduced demand for infrastructure, and enhanced safety and community vitality [112]. However, understanding how users navigate these complex environments remains a challenge, particularly in high-density, multi-level developments where both horizontal and vertical movement play significant roles.

Evidence-based designs: City planners are increasingly adopting data-driven approaches for developing smart cities [113, 114], leveraging the potential of distributed sensory networks and the Internet of Things (IoT) to gather data on urban phenomena such as traffic, energy use, waste management, and human activity [115]. At the building level, a similar concept of smart buildings is gaining traction, integrating various sensors to monitor and optimize building performance [116, 117]. In these frameworks, big data enables real-time optimization and more informed decision-making for the built environment, from street networks to neighborhood scales and down to individual buildings or even floors [118, 119, 120].

Space Syntax and Network Science: Human movement patterns within cities, streets, and buildings have been a focal point of architectural and urban studies, with methods such as space syntax frequently employed to analyze the structure of these spaces [121, 122]. However, while space syntax offers valuable insights, it is often criticized for its limitations, especially in its dual graph representation [123, 124]. Network science has emerged as a complementary or alternative approach, offering a more generalized framework to study urban systems and human interactions within them [125, 126]. The network-theoretic approach abstracts complex systems, making it particularly useful for studying multifaceted environments like integrated developments [127]. These developments, with their high density and mixed-use nature, introduce spatial and functional heterogeneity, making their analysis more complex than traditional single-use developments [128].

Sparsity of human tracking data: An essential challenge in studying integrated developments is the scarcity of suitable data to test hypotheses about human behavior within these spaces. Data-driven design is a growing

trend in urban planning based on the premise that "we cannot improve what we cannot measure"[129]. Human movement within buildings, often referred to as circulation, is a crucial aspect of architectural design, influencing factors like accessibility, safety, and connectivity[119]. Although space syntax and network theory have been applied to analyze circulation at the building level [130], studies specifically focusing on human movement in mixed-use and integrated developments remain scarce [128, 131]. Notable examples include studies that collected movement data from users in the Barbican and South Bank developments in London [128], and in mixed-use streets [131].

Data, Research, and Methodology gap: The lack of adequate movement data hinders the ability to rigorously evaluate design decisions in integrated developments. In this study, we introduce high-resolution human movement data collected in a new integrated development in Singapore, known as Kampung Admiralty (KA). This data, collected using Bluetooth beacons and mobile applications, tracked participants over several days, offering a granular view of human activity in the space. The data collection approach aligns with a network-theoretic framework, where both spatial and functional elements of the built environment are analyzed as a weighted network [127]. This methodology offers a powerful tool for studying the complex relationships between human behavior and built environments.

Our approach focuses on two spatial factors—topology and function—and their influence on user movement within KA. Function refers to the utilization of space (e.g., commercial, residential, or public use), while topology pertains to the physical layout and connectivity of spaces. The choice of Euclidean distance along the network (or routing distance) is essential for analyzing user movement in such familiar environments [132, 133]. Research shows that angular distance may better predict vehicular movement, but in pedestrian networks like KA, Euclidean distance provides a more accurate representation of user behavior, particularly for familiar users [134, 135].

By employing network representations and integrating spatial and functional factors, we provide a quantitative analysis of how building topology influences human movement. This approach allows for a more abstract understanding of movement patterns, similar to methods used in social and ecological networks. Our study highlights the intricate interplay between topology and function in a high-density, complex urban environment. Moreover, we uncover movement patterns that reflect the interplay between different spatial and functional components of the integrated development, shedding light on how these relationships shape user behavior.

In conclusion, the future of urban development, particularly in the context of integrated and mixed-use developments, necessitates a shift towards data-driven design methodologies. Such approaches can uncover the complex dynamics between human behavior, space, and function, ultimately leading to better-informed design decisions that align with the goals of sustainability, safety, and inclusivity. The KA case study, reported in Chapter 7, demonstrates the potential of network science and data-driven analysis to provide meaningful insights into human movement patterns within integrated developments, paving the way for improved urban planning practices that address the challenges of 21st-century urbanization.

2.6 Overall Conclusion

In conclusion, this chapter has established the foundational understanding necessary for the subsequent analysis of vertical and horizontal mobility patterns. It has reviewed the advancements in human activity recognition (HAR), emphasizing the role of barometric sensors in detecting vertical movements, an underexplored dimension of urban mobility. The integration of multiple sensors, including accelerometers, Wi-Fi, and barometers, was highlighted as a crucial approach to improving activity recognition accuracy, particularly in high-density, multi-level urban environments.

The chapter also traced the evolution of human mobility studies, underscoring the need to move beyond traditional two-dimensional frameworks. The concept of volumetric urbanism was introduced, advocating for cities to be studied as three-dimensional spaces, especially in light of increasing urban density and vertical growth.

Significant gaps in the existing literature were identified, particularly the lack of methodologies and large-scale datasets for analyzing vertical mobility and its interplay with horizontal movements. These gaps point to the necessity of developing comprehensive frameworks to bridge this knowledge divide.

This chapter thus provides a strong theoretical and methodological foundation for the thesis. By synthesizing insights from existing studies and identifying areas for advancement, it sets the stage for the development of a machine-learning framework to analyze vertical displacement and uncover nuanced mobility patterns, contributing to the broader field of urban mobility research

3

Data and Methods

This chapter provides a detailed account of the data sources and methodologies employed in this thesis to explore human mobility patterns, particularly in urban environments where vertical and horizontal dimensions intersect. It introduces three key datasets: Sec. 3.1: Urban mobility modes experiment data (used in Chapter 4 and 5), Sec. 3.2: National Science Experiment (NSE) 2016 data set (used in Chapter 5 and 6), and Sec. 3.3: Kampung Admiralty 2021 data set (used in Chapter 7). Each dataset was selected to address different research questions and to enable the development and validation of frameworks for mobility analysis. This chapter also contains an additional description of the data and methodology that does not fit neatly in the published (and appended) research articles, but provides more insight into collected data and analyzed results.

The chapter begins by describing the Urban Mobility Modes Experiment, a controlled study designed to capture sensor data across various modes of transportation and activity, including walking, cycling, and vertical movement such as stairs and elevators. Next, the NSE 2016 dataset, one of the largest of its kind, is introduced. This dataset tracks the daily activities of approximately 50,000 students, offering rich insights into mobility patterns on a city-wide scale. Lastly, the Kampung Admiralty dataset focuses on a

vertically integrated urban development, providing a unique opportunity to study movement in a compact, multi-functional space.

The methodologies described in this chapter include sensor deployment, data preprocessing, and machine-learning frameworks for activity recognition. Emphasis is placed on handling challenges such as noise reduction, integration of diverse sensor data streams, and addressing gaps in spatiotemporal coverage. This chapter also discusses the ethical considerations and privacy protocols adhered to during data collection and analysis. By laying the groundwork for robust data handling and novel methodologies, Chapter 3 forms the empirical backbone for the research findings presented in subsequent chapters.

Author contribution to data sets: Data set in Sec. 3.1 and Sec. 3.3 was generated with the thesis author as the lead researcher in charge of data collection. The author was not involved with the data collection experiment briefly described in Sec. 3.2, but performed all data analysis described.

3.1 Urban mobility modes experiment data

As the measure of the barometric pressure behavior is influenced by many factors, it is vital to have at our disposal a data set with the corresponding ground truth for validation purposes. Therefore, we collected approximately 6 hours of sensor data annotated using video recording for this aim. This data set was recorded across different modes of horizontal (walking, idle, train, bus, car, and cycle) and vertical (elevator, escalator, and stairs) activity (Table 3.1). Two researchers collected data on different days and times using a custom-made recording device, SENSg, hung in a lanyard (at least two devices were carried, and up to 5). There are two modes to the SENSg device, each operating in two different recording frequencies, 0.0625 Hz and 1 Hz (high frequency is only used for tests and not deployed in the field for large-scale data collection), respectively. The video was recorded using a GoPro Hero 6 mounted on the chest. We will use this data set to demonstrate the various factors influencing barometric pressure in Chapter 4 and validate manual labeling methods in Chapter 5 when identifying vertical displacement activities.

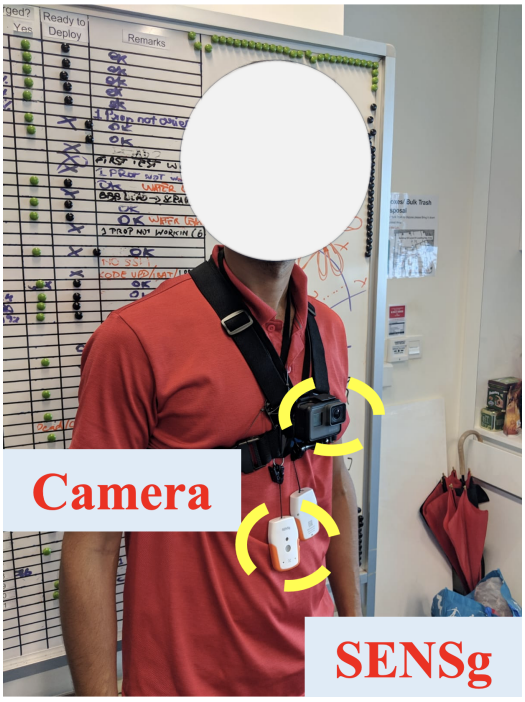


Figure 3.1: **Urban mobility modes experiment:** Experiment set-up for collecting sensor data during different mobility modes.

Table 3.1: **Video-annotated mobility data**

Mode	Total time (hours)	Additional notes
Car	0.4	Taxi trip includes passing through an underground tunnel
Bus	1.5	Two separate public bus rides with multiple bus stops and traffic stops.
Train	0.7	Train trip includes underground and above-ground stations in a single trip.
Cycle	0.8	Cycle trip includes paths with changing elevation.
Walking	1.5	Walking includes paths with changing elevation.
Idle	0.3	
Vertical mobility	0.4	Vertical mobility modes include stairs, escalators, and elevators.
Total mobility	5.6	
Total Stationary	19	(a) Devices were kept stationary for 14 hours to capture the diurnal pressure cycle in Singapore (b) 1 hour each in an air-conditioned (AC) large and small room, (c) half-hour in an AC room followed by a half hour in the non-AC room, (d) 1-hour facing wind and 1-hour behind windows (against the wind).

3.2 Singapore National Science Experiment (NSE) 2016 data

The National Science Experiment (NSE) was designed and commissioned by the Singapore National Research Foundation (NRF) and the Singapore University of Technology and Design (SUTD), with other private and government bodies in Singapore [136]. The primary objective of this island-wide science experiment carried out by Singapore students—themed "Step Out for Science"—was to monitor and evaluate their carbon footprint, travel mobility patterns, amount of time spent indoors and outdoors, and more (Fig. 3.2).

This dataset is used to derive vertical mobility activities in Chapter 5, which forms the main topic of analysis in Chapter 6, along with horizontal mobility.

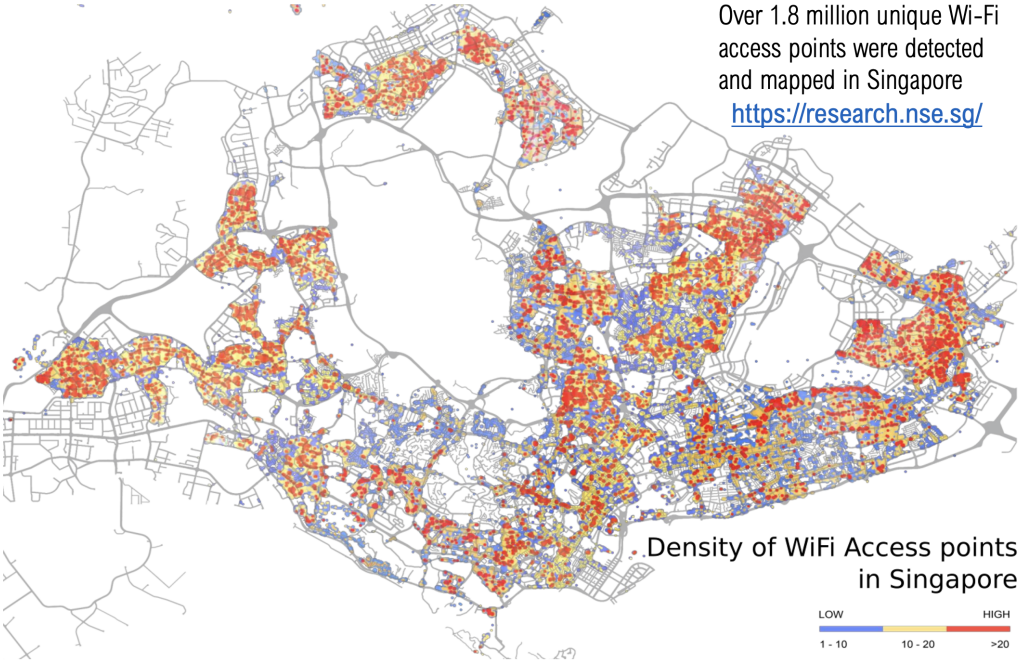


Figure 3.2: **NSE 2016 Spatial Scope:** Wi-Fi access points detected and mapped in Singapore during NSE 2016 experiment

3.2.1 Data description

Almost 50,000 students from 92 schools distributed nation-wide participated in the NSE in 2016 (Table 3.2). Each student carried a wearable device called SENSg (see Fig. 3.3), which consisted of built-in environmental, motion sensors, and communication units. The devices were able to record and transmit the sensed data related to the Activities of Daily Living (ADL) of the students to a cloud server [1]. The data was recorded every $13 \sim 18$ seconds over weekdays, from Monday to Friday, for eight weeks in 2016. The device goes to sleep mode to reduce battery consumption when the processed IMU signal shows no user movement. The SENSg devices were handed over to the students on Mondays and collected back on Fridays. Hence, the full-time scale of daily ADL is only available on Tuesdays, Wednesdays, and Thursdays. Table 3.3 shows the cleaned NSE 2016 database after removing devices based on two criteria: (a) data coverage for less than 4 hours per day and (b) percentage of missing location data larger than 30% in a day.

Table 3.2: **NSE 2016 - Device delivery information**

Dates	No. schools	No. devices	Primary (6–12 y.o.)	Secondary (12–16 y.o.)	Pre-University (16–19 y.o.)
Apr 11–15	7	3,002	1,590	1,192	220
Apr 18–22	7	3,290	2,300	990	–
May 16–20	10	2,673	820	1,624	229
May 23–27	2	840	–	220	620
July 11–15	32	10,751	3,834	6,677	240
July 18–22	23	7,934	5,233	2,181	520
July 25–29	8	15,029	220	890	13,919
Aug 17–19	3	6,000	–	–	6,000
Total	92	49,519	13,997	13,774	21,748

Sensors for detecting Vertical Mobility: Our study uses the accelerometer, barometer, and Wi-Fi scanners embedded in the SENSg devices to detect students' Vertical Displacement Activities (VDA) during their ADL. These sensor types are widely available in modern smartphones, rendering them ideal for this particular HAR. The SENSg device comprises the IMU sensor MPU9250 from InvenSense and the barometer sensor BMP280 from Bosch Sensortech. The device also collected and stored up to a maximum of 20 Wi-Fi Access Points (AP) with the highest Receiver Signal Strength Indication (RSSI). We used Skyhook, a mobile location service from Boston,

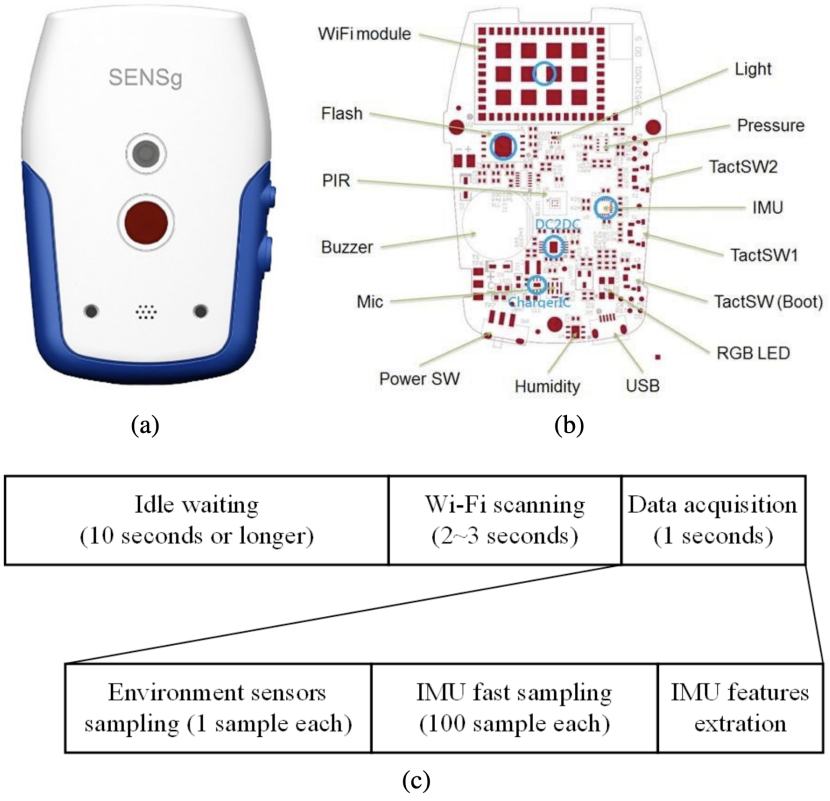


Figure 3.3: NSE SENSg device details: (a) outside look, (b) internal structure [1], (c) working cycle. (Picture Courtesy: [2])

Table 3.3: **Curated NSE 2016 database**

Dates	No. schools	No. IDs	Primary (6–12 y.o.)	Secondary (12–16 y.o.)	Pre-University (16–19 y.o.)
Apr 11–15	6	1,491	593	741	157
Apr 18–22	7	2,031	1,353	678	–
May 16–20	10	844	256	476	112
May 23–27	–	–	–	–	–
July 11–15	31	5,240	2,043	3,042	155
July 18–22	21	3,667	2,532	973	162
July 25–29	7	5,905	–	121	5,784
Aug 17–19	3	2,521	–	–	2,521
Total	85	21,699	6,777	6,031	8,891

Massachusetts, that has geolocation of billions of Wi-Fi APs around the world, to convert the Wi-Fi APs to location coordinates [137]. The location coordinates (latitude and longitude) have a typical location accuracy of ±100 meters. The location accuracy is increased by applying regression to the time-series of location data (see Appendix of Chapter 5). When the Wi-Fi APs are sparse or absent, location data is considered missing. Interpolation is then applied to the time-series data to predict these missing values (see Appendix of Chapter 5). The full sensor characteristics are shown in Table 3.4 [1]. The embedded barometer sensor is capable of detecting up to 1-meter changes in height, i.e., ±12 Pa [1]. The raw values measured by the accelerometer along its three axes were processed on-board the SENSg, and only descriptive statistics of these raw data were recorded: (1) $\max(M_{\text{acc}})$— the maximum value of the accelerometer's signal magnitude (2) $\text{std}(M_{\text{acc}})$— the standard deviation of accelerometer's magnitude, both sampled at 100 Hz during the one-second data acquisition temporal window, which occurred in its turn with frequency $\sim$ 0.0625 Hz. The SENSg device, along with its working cycle, is shown in Fig. 3.3.

Table 3.4: **Sensor Characteristics of SENSg device**

Sensor	Model	Range	Accuracy	Units	Poll frequency
Wi-Fi	SN8205	–	–	–	0.062 Hz
Barometer	BMP280	300 to 1,100 hPa	±12 Pa	Pa	0.062 Hz
Accelerometer	MPU9250	±2 g	±80 mg	mg/LSB	100 Hz (for every 0.062 Hz)

Human mobility data set: The human mobility data is derived from the data set described above and reported in Table 3.3. The definition and method of derivation of horizontal and vertical mobility are briefly noted in the subsequent sections. Students with no movement data or failed home detection are removed (5% of total). The description of the curated mobility data set is shown in Table 3.5. The NSE 2016 mobility data set contains 20,570 subjects with 318,434 vertical displacements and 249,512 horizontal displacements measured over time period $\triangle t < 3$ days. Of the 20,571 subjects, $\sim$ 7900 had mobility events for 3 days, and $\sim$ 6700 had 2 days. The data set has a relatively low time coverage as the data is recorded only when

the IMU sensor detects body movements. Nevertheless, the data set has a median time coverage of 7.2 hours daily.

Table 3.5: **NSE 2016 Mobility data set A. Time period $\triangle t < 3$ days**

Population	No. schools	No. subjects	No. VD	No. HD
All	85	20,571	318,434	249,512
Primary (6-11 y.o.)	38	6,346	98,953	93,767
Secondary (12-16 y.o.)	34	5,682	75,155	70,837
Pre-university (17-21 y.o.)	13	8,543	144,326	84,908

Vertical mobility data set for estimating building floor height The height of the vertical built environment is estimated by identifying the maximum vertical displacement in a given location (i.e., home, school, or free-range environment) based on the vertical mobility data set B (Table 3.6) that defines displacement as the cumulative distance traversed between consecutive stops without change in direction (i.e., a vertical mobility activity either contains only upward mobility or only downward mobility). Besides the difference in what constitutes a trajectory and displacement, the definition of vertical mobility in this data set in terms of waiting time ($\triangle t = 15$ minutes) and stop area ($d_{s,v} = 2$ m) is the same as that in the NSE 2016 mobility data set.

Table 3.6: **NSE 2016 Vertical Mobility data set B. Time period $\triangle t < 3$ days**

Population	No. schools	No. subjects	No. VD
All	85	20,573	389,031
Primary (6-11 y.o.)	38	6,346	115,465
Secondary (12-16 y.o.)	34	5,682	90,434
Pre-university (17-21 y.o.)	13	8,545	183,132

3.2.2 Mobility definition:

A stop S is a location within a specified area (given by diameter d_s) where a minimum amount of time is spent (known as the waiting time $\triangle t$). A trajectory T is the path travelled between two consecutive stops (Fig. 3.4).

A displacement $\triangle r$ is the total distance a trajectory covers, i.e., $\triangle r_i = \sum_{j=1}^{n} d(l_{i,j})$ where $d(l_{i,j})$ is the jump length between consecutive locations $l_{i,j}$ recorded in a trajectory T_i. The time spent in mobility is given by τ. The waiting time used in the study can be well defined as the time spent $\triangle t_i$ at the destination S_{i+1} after moving a distance $\triangle r_i$.

Vertical mobility is movement in the built environment's vertical dimension, typically through stairs, escalators, or elevators. A waiting time of $\triangle t_v = 5$ minutes within diameter $d_{s,v} = 2$ meters of vertical area defines a stop S_v in the vertical dimension (Fig. 3.4 A). Horizontal mobility is movement in the built environment's horizontal plane, i.e., across the urban environment's planar surface. A waiting time of $\triangle t_h = 15$ minutes within $d_{s,h} = 75$ meters of horizontal area defines a stop S_h in the horizontal dimension (Fig. 3.4 B). A detailed illustration and description of these three mobility types and their properties (trajectory, displacement, waiting time, mobility time) are shown in Figure 3.4.

The definition of displacements used in the study differs from some literature [8] that defines displacement as a straight line or Euclidean distance. Our data set shows a 96% Spearman correlation ($p < 10^{-6}$) between straight line (referred to in this article as origin-destination (OD) distance) and trajectory distance. Nevertheless, results generated with straight line distance are shown in the supplementary information file (attached as an appendix in Chapter 7).

Choosing mobility parameters: The choice of waiting time $\triangle t$ should reasonably reflect the mobility behavior of humans. For example, we define displacement as the distance a user travels when motivated to move from one place to another. While assessing what motivates an individual is complex, the stop point should ideally ignore re-fuelling, transit changes, or traffic stops that are not directly relevant to what makes a user move from one place to another.

In introducing mobility in the vertical dimension, we must consider an appropriate waiting time $\triangle t_v$ different from the factors influencing our choice for $\triangle t_h$. For example, some key factors to consider are (a) the spatiotemporal scale of vertical mobility is much smaller, (b) transit time between vertical transportation modes such as a lift is also typically smaller than the transit time between two buses or trains, and finally (c) motivated destination within a building or a development complex is also plenty and diverse than in the horizontal dimension at the urban scale [11]. For these reasons, the $\triangle t_v$ is chosen to be smaller than $\triangle t_h$. Results generated with equal wait times

($\triangle t_v = \triangle t_h = 15$ minutes) are shown in the supplementary information file (attached as an appendix in Chapter 7). The choice of stop radius is informed by the minimum spatial and temporal resolution available in the NSE 2016 data set (discussed later in the section).

Our mobility data set does not sufficiently represent movement within $\pm d_s/2$ meters of a stop, as is typically expected by choice of d_s. Results generated with different waiting times ($\triangle t_v = 10$ and 15 minutes and $\triangle t_h =$ 10 and 20 minutes) are shown in the supplementary information file (attached as an appendix in Chapter 7).

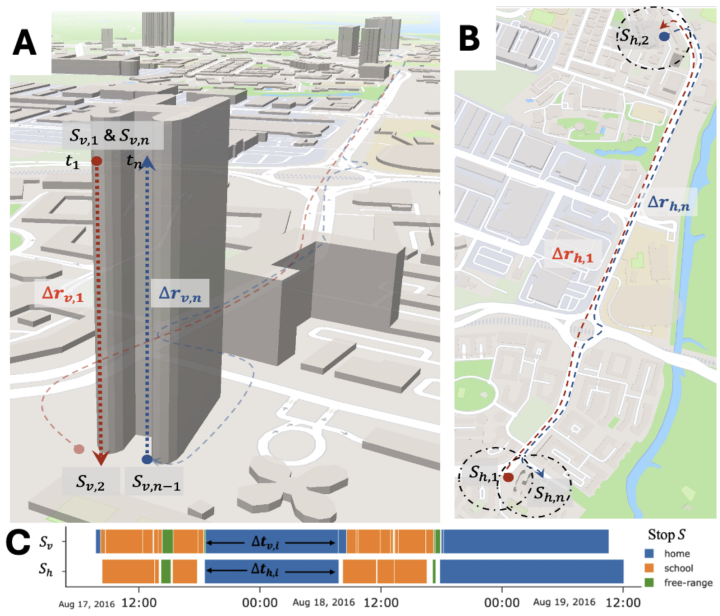


Figure 3.4: **Mobility illustration and temporal patterns A,** Illustration of vertical displacement Δr_v and stop S_v along the vertical plane. The red dotted line with a downward pointing arrow in the vertical plane of the building indicates the downward vertical trajectory $T_{v,1}$ characterized by vertical displacement $\triangle r_{v,1}$ with elapsed (mobility) time $\tau_{v,1}$. The waiting time after moving vertically $\triangle r_{v,1}$ is given by $\triangle t_{v,1}$ (it is the time spent at the stop in the vertical plane $S_{v,2}$). The vertical plane of stop $S_{v,2}$ extends along the city's surface until moved again in the vertical plane. **B,** Illustration of horizontal displacement Δr_h and stop S_h along the horizontal plane. The black dot-dash lines indicate the diameter of the stop area $d_{s,h}$. **C** Timeline of movement and stops for a sample subject. White block represents movement. Vertical and horizontal movement is observed to take place very close to each other.

3.2.3 Derivation of vertical human mobility

To the best of our knowledge, the derivation of vertical human mobility has never been done at the city scale before. The geographic (city) and temporal scale (few days) provide unique challenges. As such, new frameworks were developed to tackle such challenges by drawing knowledge and methodology from various fields of science and engineering.

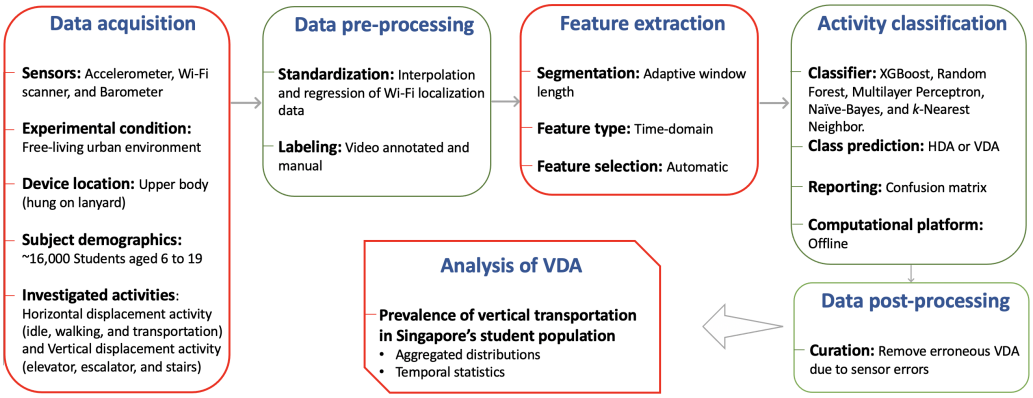


Figure 3.5: Human activity recognition (HAR) process in the identification of Vertical Displacement Activity (VDA) [3]. This flowchart describes the overall process developed to achieve our VDA analysis, which is represented in the last rectangular box. The key contributions of this work are in the boxes with a red outline.

Vertical mobility is derived using the machine learning framework developed and reported in Chapter 5 [10], based on a comprehensive understanding of barometric pressure behavior reviewed in Chapter 4 [9]. An overview of the machine learning framework to derive vertical human mobility is shown in Fig. 3.5. It consists of the following steps—data acquisition, data pre-processing, feature extraction, activity classification, and data post-processing. Using data from the accelerometer, Wi-Fi-based localization, and barometer, the machine learning framework provided 98% overall accuracy and 92% F_1 score in classifying vertical mobility [10]. Furthermore, post-processing removes vertical displacement around high-pressure fluctuation regions ($\sim$ 8% of total) for a conservative estimate. The change in barometric pressure data identified during a vertical mobility event led us to calculate the vertical displacement. The barometer sensor detects vertical displacement with an accuracy of ± 1 meters. However, due to the relatively

low sampling frequency of 1/16 Hz, the minimum vertical displacement identified is 2 meters. A barometric pressure change $\triangle P$ of 12.01725 Pascals corresponds to a vertical displacement $\triangle r$ magnitude of 1 meter (accurate for $\triangle r < 10$ km from sea-level [30]).

3.2.4 Derivation of horizontal human mobility

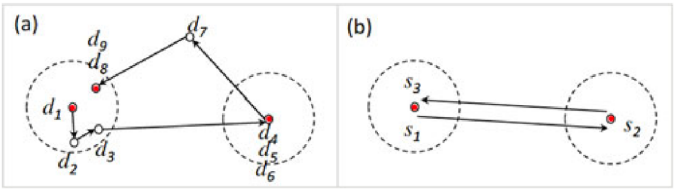
Stops are identified using a standard stop-point detection algorithm using the Python library Moving Pandas [138]. The stop detection algorithm detects a stop if a user has moved within an area (given by a diameter $d_{s,h}$) and spent a minimum period (waiting time $\triangle t_h$). The parameters used in the main article are $d_{s,h} = 75$ meters and $\triangle t_h = 15$ minutes. Due to relatively low spatial resolution, the minimum horizontal displacement recognized is 20 meters. The horizontal displacement $\triangle r_{h,i}$ is the sum of the Haversine distance (i.e., great circle distance) between the time series location data in a given trajectory $T_{h,i}$.

The diameter of the stop area $r_{s,h}$ is informed by the spatial resolution of the Wi-Fi-based localization. The original location data has an accuracy of < 200 meters for 90% of the data (99% data within < 300 meters). Location data has low accuracy but relatively high precision (91% of consecutive locations have jump lengths < 50 meters). However, low accuracy can overestimate the trajectory distance. Hence, the time series location data is then smoothed using the Kalman Filter (KF) that employs a nearly-constant velocity (CV) model to reduce the noise introduced by low spatial resolution (98% of data now has a precision of < 100 meters and 96% has < 50 meters). Using the Moving Pandas python library's KF CV smoother [138], data is smoothed with parameter values (a) process noise uncertainty factor of 0.1 and (b) measurement noise error factor of 50. The results generated with stop diameter $d_{s,h} = 50$ and 100 meters are shown in the supplementary information file (attached as an appendix in Chapter 7).

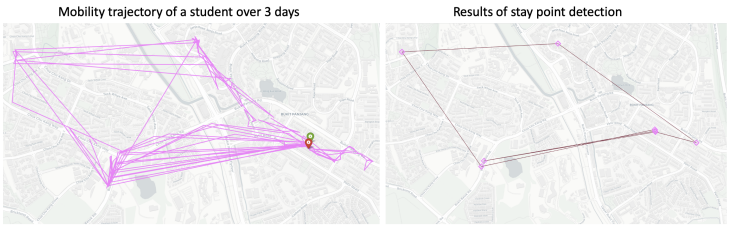
3.2.5 Home and school detection

School names and locations for each subject are known factors. The mobility data within a school is detected by identifying location data within school zones. School zones are given by two sources: (a) school zone polygons provided by the Singapore government (data source) and (b) drawing custom square grid around the known school location with a radius of 150 meters taking into account the systematic deviation of location data (identified visually by plotting all the location data from the NSE 2016 data set).

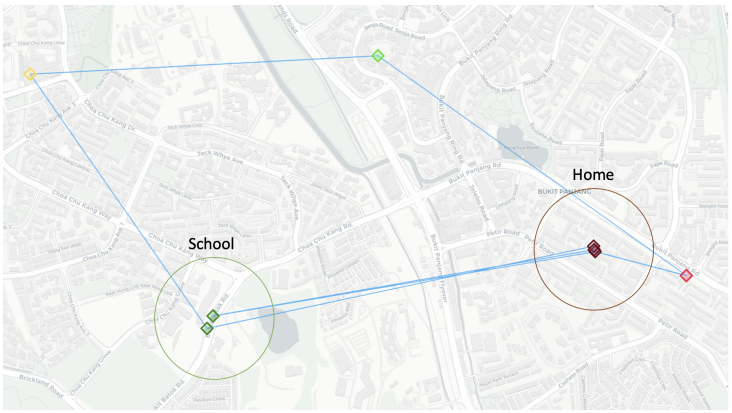
Home detection involves identifying where users spend most of their nighttime based on the last detected stop in a given day (Fig. 3.6c). A custom square grid with a radius of 200 meters from the detected home location is used to identify displacements that occurred or originated in the home.



(a) Stop/Stay point detection. d denotes trajectory points and s indicates stop points [59]



(b) From trajectory to stop points



(c) Home and school detection

Figure 3.6: Stops, trajectories, and place detection.

3.2.6 Empirical distribution fit

The displacement and waiting time distributions have been found to have one or more (multi-scale) of the following heavy-tailed distributions: power law, truncated power law, Log-normal, Stretched exponential, and Exponential distribution [77]. All of these distributions are considered plausible candidates to fit our empirical observations. The following are the standard steps taken to find the best fit and respective parameters (adopted from Barabasi [126] and Alessandretti et al. [77]).

- Step 1: Choose the minimum (x_{min}) and maximum value (x_{max}) of the given data for which the fit is to be applied. The choice of x_{min} can start with the known minimal resolution of the data. For example, the minimum spatial resolution of vertical displacement in the NSE 2016 data set is 2 meters. However, the values close to 2 meters may not be well represented in the data set due to low temporal resolution. Hence, the choice of x_{min} starts from 2 meters and increases till an optimal fit (see next steps) is obtained. However, x_{min} cannot be infinitely increased and needs to be capped such that most of the data is preserved, even if it is not the optimal fit.

- Step 2: For each x_{min} and/or x_{max} and given distribution type, the respective parameters that best fit the empirical data are found by Maximum Likelihood Estimate (MLE) i.e., fit parameters that maximize the log-likelihood function.

- Step 3: Kolmogorov-Smirnov (KS) statistics (or simply distance D) is then calculated to measure the distance between the synthetic or theoretical fit and the empirical data. It is used to find the optimal x_{min} and x_{max} values. It is given by,

$$KS = max(|F - P|), \tag{3.1}$$

  where F is the cumulative distribution of the best fit, and P is the cumulative distribution of the empirical or synthetic data.

- Step 4: The optimal x_{min} and/or x_{max} is the fit that minimizes D. Specifically, the optimal values are obtained by the elbow method, where the D crosses the threshold (critical junction) of rapid decrease to slow descend or stead-state, forming the elbow-like shape. In cases where this trend is unclear, we choose the x_{min} that removes less data.

- Steps 2 to 4 are repeated with different x_{min} and x_{min} values, where one is fixed at a time while varying the other until optimal x_{min} and x_{min} are found. Step 2 is repeated for the last time with optimal x_{min} and x_{min}.

- Step 5: Once the optimal distribution parameters are found, the candidate distributions are compared to find the best fit. The log-likelihood and the number of free parameters for the distributions are used to calculate the Akaike Information Criterion (AIC). The best fit among the candidates is the one with the maximum AIC weight (AIC^w) calculated from the relative log-likelihood of the candidate fits.

  For a distribution model m,

  $$AIC_m = -2 \log L_m + 2V_m + \frac{2V_m(V_m + 1)}{n - V_m - 1}, \tag{3.2}$$

  where L_m is the maximum likelihood for the candidate model m, V_m is the number of free parameters of the model, and n is the number of sample data.

  $$AIC^w = \frac{e^{-0.5(AIC_m - AIC_{min})}}{\sum_{k=1}^{K} e^{-0.5(AIC_k - AIC_{min})}}, \tag{3.3}$$

  where K is the number of candidate models and AIC_{min} is the minimal AIC score found among the candidates.

- **Robustness of the fit:** By bootstrapping data 1000 times for 100 subjects data, we find if the candidate model that fits the overall empirical data is also the best fit for its sub-samples. We also measure the uncertainty of the fit parameters and present it through its mean and standard deviation, i.e., $\mu \pm \sigma$.

It is important to note that these steps, while widely followed, have limitations (refer to Chapter 4 of Barabasi [126] for a detailed discussion). This includes the complexity of working with real-world data. The derived exponents and fitting distributions are documented to vary widely with different data sets [77]. Hence, our primary focus will be determining if the distributions under study are heavy-tailed (implying heterogeneous behavior) and the degree of heterogeneity (power-law-like or exponential-like) and not on

the precise implications of the best candidate model (which can reveal the type of underlying process that drives human mobility behavior) or their exponents.

3.2.7 Robustness and uncertainty

All key values are represented as median and median absolute deviation ($\tilde{x} \pm MAD$) or mean and standard deviation ($\mu \pm \sigma$). In appropriate places, $\mu \pm \sigma$ is based on bootstrap values where 100 subjects' data is randomly drawn 1000 times. The following variability tests are conducted and included in the SI file (attached as an appendix of Chapter 7). Pause time variation: $\triangle t_v = 5$, 10, and 15 mins, and $\triangle t_h = 10$, 15, and 20 mins, stop area diameter variation: 50, 75, and 100 m, and displacement definition variation: Trajectory distance vs. straight-line/Origin-Destination OD distance.

3.2.8 Data and Privacy statement

The access to the NSE 2016 data set was provided by SUTD, Singapore (data hosting institution) and NRF, Singapore. Individual data were stored under anonymized identities, with no directly identifiable information. However, approximate home location (within 200 meters) and school information can be inferred indirectly after data processing. For these reasons, only a restricted form of the data will be publicly accessible.

The authors did not participate in the data collection; only the anonymized or aggregated form of the data was used for the study presented here. All authors who worked with the data set obtained necessary permission from the hosting institution and adhered to established rules.

3.3 Kampung Admiralty (KA) experiment 2021 data

The Kampung Admiralty (KA) 2021 experiment was designed as a pilot study to test the effectiveness of data-driven and Complexity Science based approach (specifically, Network Science) to evaluating urban built environment design. The experiment was conducted by the Singapore University of Technology and Design (SUTD) researchers and commissioned by the Singapore government agencies, including the Ministry of National Development (MND), Centre for Liveable Cities (CLC), Housing Development Board (HDB), and others.

'Kampung' means 'village' in Malay. Kampung Admiralty is thus an integrated development designed to serve multiple functions (residence, hospital, shopping center, and other public facilities) that we usually see distributed in a neighborhood or a city. This allows us to test the network-science-based and evidence-based approach to evaluating high-density built environments on a manageable scale. This thesis uses the KA development and KA 2021 data set consisting of human tracking data (mobility) in Chapter 7 as a case study to examine the factors influencing user movement (three-dimensional) in a vertical, high-density environment. The following sections describe the data collection system and processing, movement data, study site, and the derived spatial network.

3.3.1 Data description

The data set consists of movement data collected from regular users of the Kampung Admiralty building with duration of recording per participant varying from minimum one day up to a month in February 2021. Volunteers were recruited by representatives in booths or through posters placed around KA. They were offered vouchers worth S$30 in two installments—during recruitment and later at the end of the data collection effort—as an incentive to keep the mobile app installed. Each participant is tracked continuously for the entire duration of the experiment (i.e., day and night). However, their location can only be determined if they are ambulating within KA. To reduce the impact of a few users with long tracking data from dominating the movement data, the maximum number of days per user is restricted to 5 days (refer Fig. 3.8a for a detailed breakdown of the number of days per participant). The group of 51 participants comprises 21 males and 30 females

Figure 3.7: Aerial view of Kampung Admiralty with its tiered urban spaces at elevated levels (Source: WOHA architects, Photo by Patrick Bingham Hall)

between the age of 17 up to 90 years old, with 50% below 59 years old (refer Fig. 3.8b for a detailed breakdown of participant age). The high elderly demographic is representative of KA's regular users whose residential complex is meant to house seniors and includes community facilities like Active Ageing Hub. Of the 51 participants, 18 are residents, 9 are employed within KA, and 24 are frequent visitors. A total of 131 participating days (i.e., sum of 1-day worth of data for all users) from 51 participants were extracted from the KA data set.

Derivation of three-dimensional mobility in KA: The trip data is derived from the time series of predicted locations using the transformation table. Each trip data is a path through the KA spatial network. This also allows us to fill in missing nodes (location) in a given path by reconstructing it using the shortest path between the known locations for a conservative estimate. The predicted series of nodes in a path has a length of 3 nodes or less for 87% of the missing patches applied. Hence, even though this trip prediction method partly depends on the topology (shortest path calcula-

tion), the above numbers suggest that it does not significantly impact our study on the correlation between topology and user movement. A total of 581 paths were derived, of which 361 are continuous in time with a pause time of fewer than five minutes at each stop (i.e., ≤ 5 minutes). These 361 paths are defined as a trip in this study, whose origin and destination form the entries of the Origin-Destination matrix. The derived 581 paths encompass 510 paths from weekdays and 71 paths from weekends. A brief comparison between weekdays and weekends is shown in Supplementary Fig. S8 and S9, and Supplementary Table S5 and S6 (refer to Chapter 7). It is important to note that the data obtained on weekends is too small to make a meaningful comparison.

A pause of 5 minutes is selected based on the following consideration: (a) a pause should reflect a destination with sufficient time to perform activities in KA such as visiting a store, or making inquiries in community facilities, (b) allow minimal time for users to move from one location to another, (c) account for wait times in lift lobbies or potentially missed localization. In human mobility research, a pause time of $10 \sim 20$ minutes is considered typical to define a trip for traveling across cities or large geographic areas. [77]. However, KA's compact space and multiple opportunities for destinations require a relatively smaller pause time. Please refer to Supplementary Fig. S2-S7 and Supplementary Table S1-S4 online for results obtained with a pause time of 10 minutes (refer to Chapter 7). More information on the study site and the KA spatial network is given in the following sections.

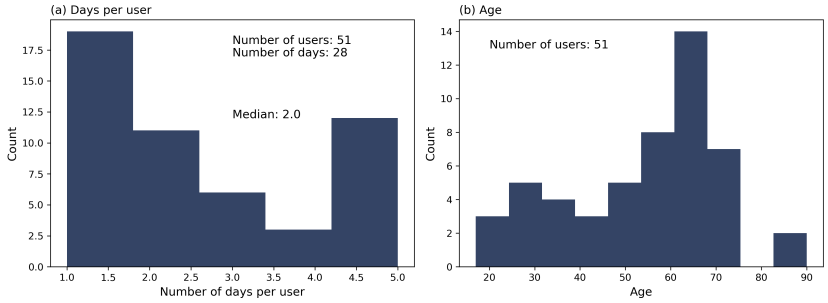


Figure 3.8: **KA Movement data—Histogram of** (a) number of days per user and (b) Age.

3.3.2 KA experiment design and set-up

The data collection system is based on a peer-to-environment Bluetooth localization method that consists of three components: (1) stationary low-energy Bluetooth beacons (Kontakt Smart Beacon Pro BP16-3), (2) custom-built data collection mobile app (iOS and Android), and (3) an access to the AWS cloud server (Fig. 3.9). One hundred twenty-four beacons are placed in locations of interest within and around the KA building perimeter. They transmit their unique ID with a sampling interval of 10 Hz, which is scanned and stored by the participant's smartphone, and the information is eventually uploaded to the cloud server.

Beacon signal range test experiment To test the signal range of the Bluetooth beacons, an experiment was devised to test for different conditions – beacon facing direction (0, 90, and 180 degree from the mobile), mobile screen lock (on and off), distance of beacon from mobile (0.5 to 50.5 meters), and transmission level (1, 2, and 3). In an open ground, the beacons were mounted on a pole and placed at 1 meter interval from each other for the first 5 meters (5 beacons), and the interval is increased to 5 meters for the next 45 meters (10 beacons)(refer to Fig. 3.10). Two mobiles (Samsung Note 8 and Google Pixel 4a) were placed 0.5 meters from the first beacon to scan and collect Bluetooth frames using the Android mobile app. The mobile app collected data for 5 minutes each in different experimental set-up.

The results show that the screen lock-on condition records data at much lower sampling rate than screen lock-off condition (later, our custom-built mobile app was designed to solve this, and allowed us to track as long as the mobile detects movement even if the screen lock was on). Other experimental conditions were tested with mobile app open on screen with screen-lock off to provide the best scan results possible. When beacons were facing the mobile at 0 degrees, signal from beacons placed till 50 meters were scanned, but at 90 degree only beacons placed till 30 meters were scanned, and for 180 degree (beacons faced against the mobile), beacons beyond 15 meters were not detected. For the experiment with different transmission level, the beacons were facing 0 degrees from the mobile and their signal frame configuration were changed from level-1 (low range) to level-3 (high range). The results showed that the beacons were scanned till 50 meters for level-3 (Fig. 3.11), 30 meters for level-2, and up to 15 meters for level-1 (Fig. 3.12).

The test results from this experiment for different transmission levels were used to find the optimal placement of beacons in Kampung Admiralty (refer

to SI file attached to Chapter 7) such that there are minimal number of beacons placed in large zones, and minimal signal interference in small and dense zones.

Mobile app testing and development: A custom mobile application was developed by a third-party company, in colloboration with the authors (assisting with overall architecture design and testing). It was designed to collect tracking data (beacon signal) in all real-world conditions.

To test the mobile app, following experimental conditions are tested–(a) mobile screen lock is on or off, (b) mobile is under movement and rest, (c) mobile app can scan one or more beacons nearby. Figure 3.13 shows the experimental set up where the mobile is placed in a toy train to simulate movement for long periods of time while being exposed to the same set of beacons. The experiment successfully showed that the app recorded beacon data in all established conditions and allowed us to deploy it for the KA experiment. The mobile app was further tested in KA under real-world conditions and this data is used as training data for a machine learning framework (refer to next section). The mobile app was made available to download from the Android and IOS markets.

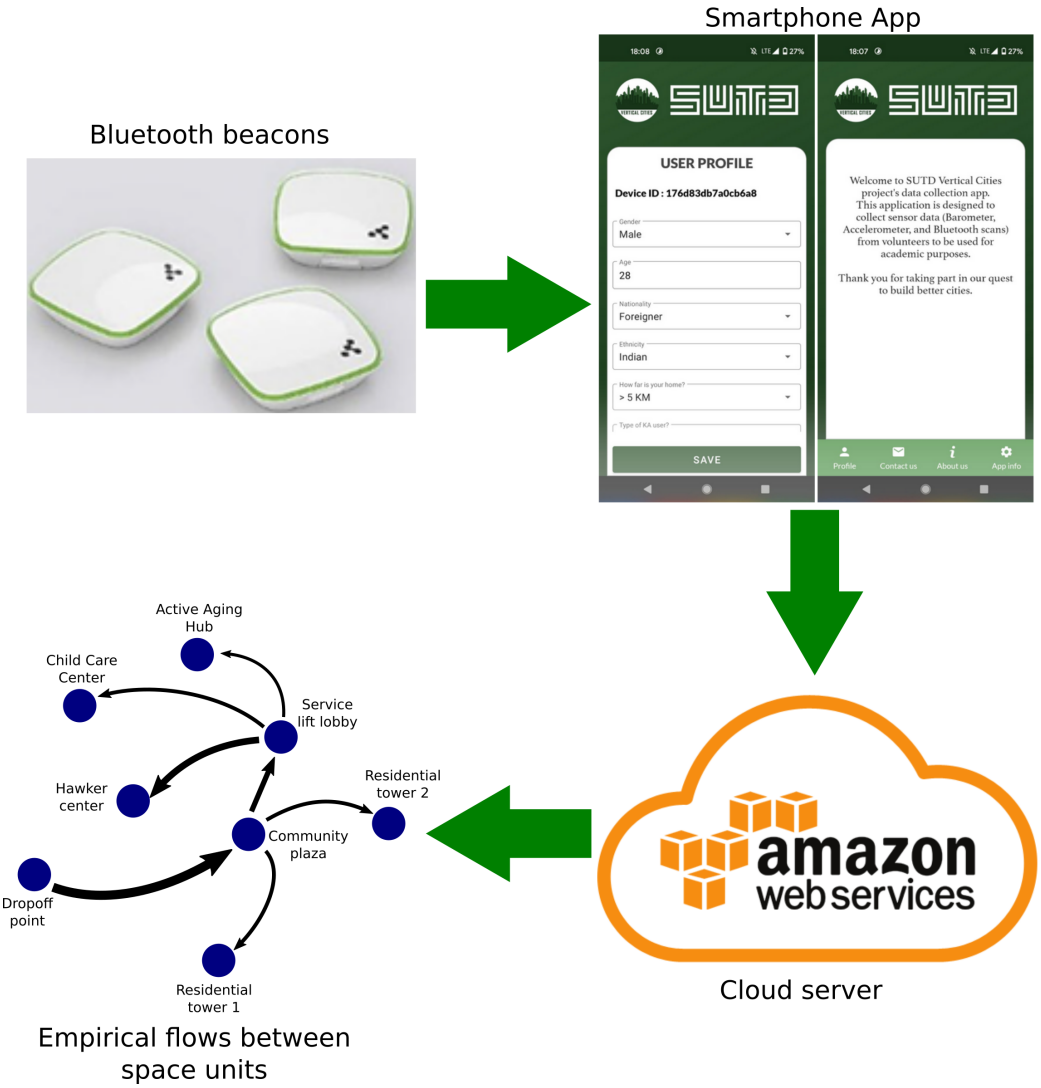


Figure 3.9: The data collection framework: the beacons were installed at various places in KA; the participants' smartphone App would scan, record, and send the nearby beacons' information to the Amazon cloud server; the data from the Amazon cloud server was downloaded and processed to represent the the flows of participants between nodes

Figure 3.10: Experimental set-up of Beacon signal range test. In an open ground, each beacon is mounted on a pole and placed at 1 meter interval from each other for the first 5 meters (5 beacons), and the interval is increased to 5 meters for the next 45 meters (10 beacons). The first beacon is placed 0.5 meter from the mobile and the last beacon at 50.5 meters

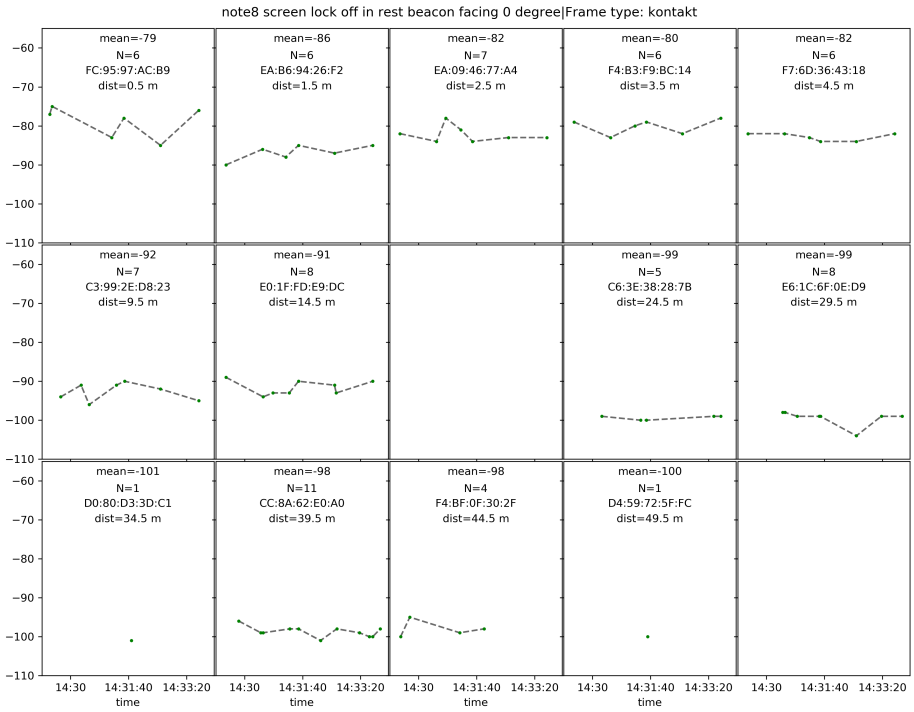


Figure 3.11: Beacon signal range test using Android app. Mobile set-up: At rest with screen lock-off; Transmission level (TX): 3; Mobile model: Samsung Note 8. TX-3 allows the beacon to transmit signals for long range (50 meters).

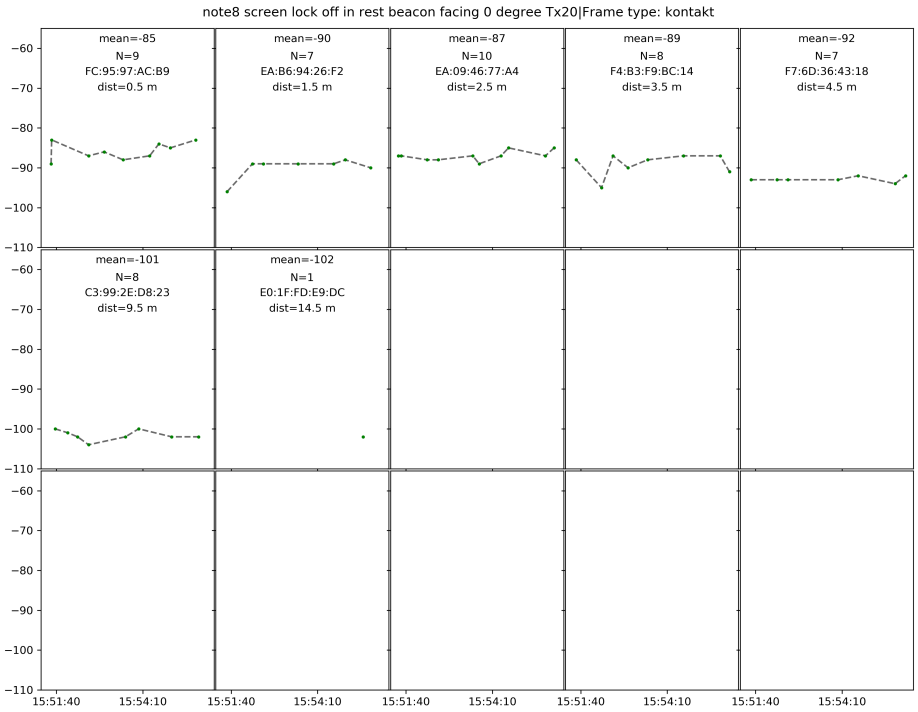


Figure 3.12: Beacon range test using Android app. Mobile set-up: At rest with screen lock-off; Transmission level (TX): 1; Mobile model: Samsung Note 8. TX-1 allows the beacon to transmit signals for short range (15 meters).

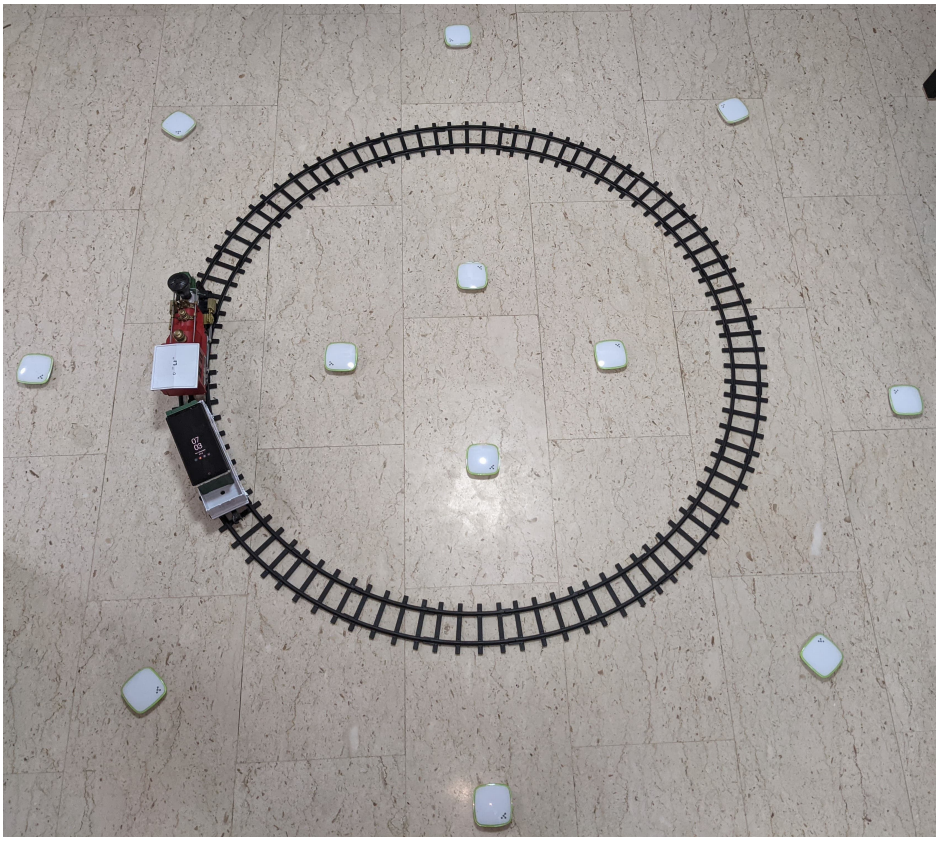


Figure 3.13: Experimental set up for mobile App testing with beacons. This allows us to record the same set of beacons for long periods of time while keeping the mobile in motion

3.3.3 Identify a user's location in KA

The beacon placement strategy is based on the varying sizes (and signal-blocking physical features) of the location grid, which calls for the customization of the beacon transmission range ($10 \sim 50$ meters). That also allowed us to reduce the signal interference between the beacons and limit the number of beacons detected at each location grid to $1 \sim 3$ units by design. In turn, this enabled us to implement a straightforward and accurate localization method within the building. This localization is based on a simple transformation table that relates the scanned beacon with the highest Receiver Signal Strength Indicator (RSSI) to a single location grid/zone. To create this table, four hours of training data were collected using two smartphones (Google Pixel 4a and Samsung Note 8). Two researchers carrying these mobile devices acquired this ground-truth data while covering the process area of KA systematically, recording the data with a sampling rate of ~ 0.2 Hz.

The exact same path was recorded twice with the two smartphones. Subsequently, a Naive-Bayes machine learning model is trained on 60% of the training data collected from one mobile and tested on the remaining 40% collected from the second mobile phone. The model achieved a classification performance of 100% accuracy at floor level and 80% accuracy at the node level. Furthermore, it reached a 93% accuracy at a slightly coarser location prediction involving the nearest neighbor. The nearest-neighbor level prediction accuracy of 93% is suitable for our study as our analysis consists of grouping the individual location zones and studying the movement data by aggregation. The transformation table that maps the beacons (with the highest RSSI) to each location grid/zone is thus generated and applied to the mobility data set.

3.3.4 Study site: Kampung Admiralty

Kampung Admiralty is Singapore's first vertically integrated public development (11 stories and 2 levels of basement, completed in May 2017) that brings together a mix of public facilities and services under one roof (see Fig. 3.14) As the increasing urban density in Singapore demands creative ways of intensifying land-use effectively in the vertical dimension, the elevated and layered urban design and architecture approach to the project led to transforming the 0.9 ha site into a dynamic vertically integrated mini-neighborhood for the community. KA's designers used a 'sandwich layered' approach to vertically join various urban functions into that 0.9 ha area.

This enhanced the mixed land-use management and living experience of the residents of KA and its vicinity. As an integrated development, the KA is immediately adjacent to a bus stop and a train station (mass rapid transit: MRT station Admiralty). We consider KA to be a high-density development due to the following reasons—(1) population density: KA is placed in a residential zone, the Woodlands East sub-zone, where the population density is very high ($\sim 38'000/km^2$), (2) Floor Area Ratio (FAR): KA's FAR is 3.6, a relatively high ratio for a residential zone, and (3) integration of multiple functions: KA integrates various amenities and services within a relatively compact area, a characteristic of high-density developments, aiming to optimize land use and promote a sense of community.

KA consists of two main parts: the community side and the residential side. The latter comprises two residential towers (the plus/cross shape towers in Fig. 3.15). The rest of the area forms the community side, which includes multiple types of facilities such as childcare, elderly care, car-park, hospital, shops, bank, and restaurants (refer to Supplementary Fig. S11 online for the distribution of these functions across the floors).

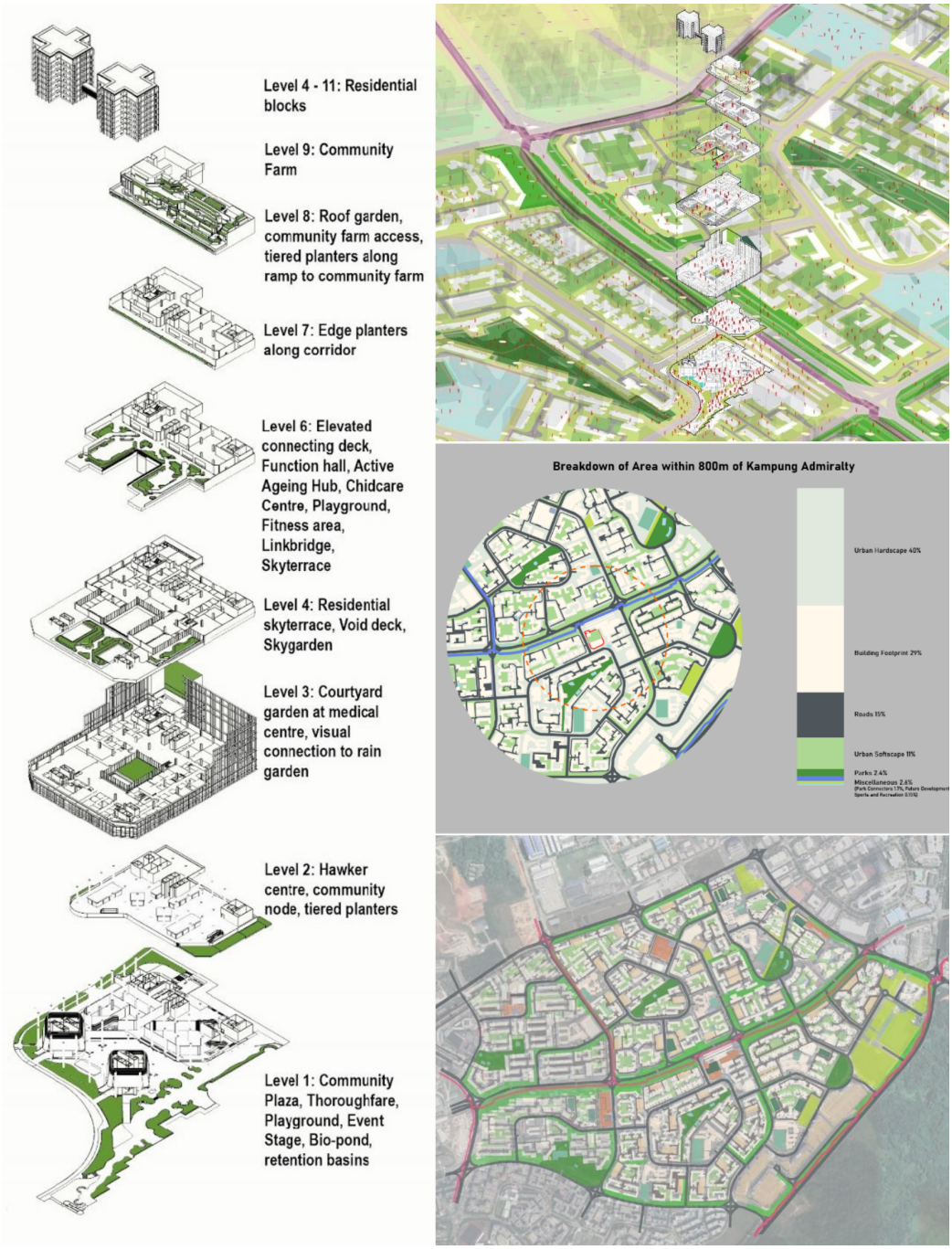


Figure 3.14: Kampung Admiralty (a) Explored Axonometric view of KA (b) KA's neighbourhood) (Reproduced with permission from Daniel Wong)

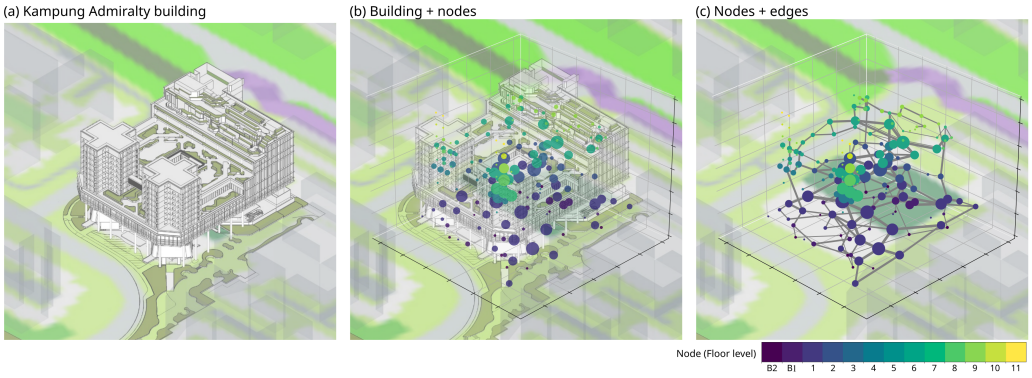


Figure 3.15: **The study site.** (a) The Kampung Admiralty building, (b) building and nodes, where size indicates incoming flows and color indicates floor level, (c) nodes and edges (weighted by the sum of flows in both directions).

3.3.5 KA spatial network

KA provides a mix of public and common spaces distributed across multiple levels and a mix of spaces with recreational facilities. The whole public space of KA is subdivided into a total of 165 space units (Fig. 3.15(b)), and the adjacency connections are linked as shown by the edges (Fig. 3.15(c)). A *node* in the KA network is defined as a programmed space with defined boundaries. An *edge* of the network is directed, and it is formed between two reachable adjacent spaces. Thus, the KA spatial network has 165 nodes and 476 edges (Table 3.7). The routing distance between the centroid of any two nodes is assigned as an edge weight. The KA spatial network consists only of public, open, and freely accessible spaces as nodes. It is primarily built to analyze the relationship between public spaces and their usage based on the participant tracking data. Many closed or private spaces like shops, restaurants, medical facilities, etc., are not considered here as nodes. The exploded axonometric view of KA with functions in each floor and the abstract representation of nodes (and edges) in each floor are shown in Supplementary Fig. S11 and S12 online respectively (refer to Chapter 7). The nodes are manually determined from the floor plans and the routing distance between them is measured from KA's floor plans. The network is constructed and analyzed through the NetworkX library in Python [139].

The aggregated movement flows are shown in Figs. 3.15(b) and (c), in which the amount of traffic arriving at a node is shown by the node size, and the size of the edges measures the total edge weight irrespective of the flow direction.

Aggregated spatial networks

The movement data is also studied on an aggregated network. Two types of movement data are aggregated : (a) user movement from one location to its neighboring location in the network (4,655 adjacent movements), (b) origin and destination of a trip taken by a user continuously moving with a pause time of less than five minutes in each location (361 trips). By aggregating the data by some combination of floor levels, we preserve the spatial adjacency and aim to study the movement flow across the vertical dimension. The floor-group components of the spatial network is also studied as sub-graphs without aggregation, thus preserving the individual location nodes and used to study within-group movement flows. The movement data is also aggregated by function that ignores the spatial adjacency and used to understand

Table 3.7: **Basic information about KA's spatial adjacency network, including number of nodes (N), number of edges (E) within floor-group, and the mobility of nodes (incoming flow) and edge.**

Floor-group	N	E	Total flow	Node's flow range	Node's flow mean (std)	Edge's flow range	Edge's flow mean (std)
All nodes	165	476	4655	0–154	28.21 (±37.03)	0–67	9.78 (±13.47)
Car-parks	25	66	183	0–42	7.32 (±9.75)	0–24	2.77 (±5.12)
Level-1	33	102	1450	0–140	43.94 (±37.27)	0–67	14.22 (±14.35)
Level-2	19	54	741	0–144	39.00 (±43.86)	0–64	13.72 (±15.79)
Level-6*	25	72	878	4–131	35.12 (±32.25)	0–45	12.19 (±12.57)
Rooftop garden	19	46	77	0–11	4.05 (±3.50)	0–6	1.67 (±1.85)
Residential*	44	88	890	0–129	20.23 (±32.20)	0–55	10.11 (±15.62)
Other	9	16	61	0 – 40	6.78 (±12.97)	0 – 40	3.81 (±10.29)

*Level-6 and residential towers share several nodes on the sixth floor where they are interconnected, including the connection bridge and several public spaces. Category Other contains L3 (1 node), L4 (1 node), and L7 (7 nodes).

the movement flow between function categories.

Aggregation by floor: In this study, we adopted a specific framework to represent the multiple floors of this elevated building. Basements (B1 and B2) are grouped and defined as car-parks. Each of the levels L1, L2, and L6 became a group. Level-1 is the ground floor of KA, which is afforded a large open plaza space, several shops, and some entrances from different directions in the neighborhood. Level-2 is a food court, which can only be accessed by walking from L1 or the community lift connecting to the rest of KA's community side. Level-6 includes a childcare center, an active ageing hub, a playground, a fitness corner, a garden, and two connections to the social spaces in the residential tower side. Level-6 of the residential towers contains some public social spaces with two connections to the community side of KA that shares the space with the rest of L6. Therefore, the L6 floor and the residential towers share those public social spaces in the network analysis. Level-8 and L9 have a community farm, a community garden, and several sitting places; the two floors were grouped as a rooftop garden. Level-3 and L4 of KA is a medical center, which can be accessed by appointment only, hence containing only one node (lift lobby) for each of these two floors. In L7, the two community facilities (childcare center and active ageing hub) are connected to the lift lobbies, making seven nodes. These three levels (nine nodes) are grouped as 'other' and would not be discussed due to their simple structure. Some basic statistics of the network and the six floor groups (car-parks, L1, L2, L6, rooftop garden, and residential) are shown in Table 3.7.

Aggregation by function: The network is also aggregated by function (Table 3.8). A total of eight function categories are included: (a) entrance (12 nodes), (b) social (15 nodes), (c) commercial (24 nodes), (d) community (18 nodes), (e) garden (23 nodes), (f) residential (14 nodes), (g) vertical (48 nodes), and (h) corridor (11 nodes). The analyses and discussions in this study used both floor and function categories as the grouping methods.

3.3.6 Data and Privacy statement

The generated network and relevant materials are available at the following public GitHub repository (`https://github.com/ajaymanivannan/Kampung_Admiralty_Dataset_v2`), which includes the spatial network file with mobility data, anonymized participant's trip data, and beacon to zone transformation table.

The research involving human subjects and the respective experiment protocol has been reviewed and approved by the Singapore University of

Table 3.8: **Spatial functions**

Function	Locations
Entrance	Dropping points and walkways connecting to KA
Social	Seating area, community plaza, fitness corner
Commercial	Restaurant, bank, food court, supermarket, shops
Community	Childcare, elderly care, car-park
Garden	Garden around sitting area, walkways, and roof top gardens
Residential	Apartments
Vertical	Lift, escalator, and stairs lobby
Corridor	Corridor areas not associated with above

Technology and Design (SUTD) Institutional Review Board (IRB) under reference code S-20-353 (the institution under which the experiment was conducted). Informed consent was obtained from all human participants, and the experiment followed the necessary guidelines and regulations concerning research involving human subjects.

In addition, the author of this thesis took mandatory courses to conduct social and behavioral research developed by the Collaborative Institutional Training Initiative (CITI Program). The author's certificate can be verified by the following link (`https://www.citiprogram.org/verify/?k4b100df3-daf2-4f26-99dc-497e42e94d6e-38055189`).

3.4 Technical tools and software

All articles, including this thesis, were written using Overleaf (LaTeX).

All computing scripts were written in Python. Database operations were performed using the Postgresql Python library. Main Python libraries used include Networkx (network science-related works), Moving Pandas (mobility-related algorithms), Scikit-learn (machine learning), Numpy, Scipy, Math, and other dependent libraries. Specialized library use is mentioned in the respective articles, and in most cases, the corresponding publications are referenced accordingly. Networks were visualized and analyzed partly using Gephi software.

Part II

How to derive Vertical Mobility?

4

On the Challenges and Potential of Using Barometric Sensors to Track Human Activity

This chapter includes a peer-reviewed paper published in MDPI Sensors, which addresses the first research question (R1) posed in Chapter 1, focusing on the factors that influence barometric pressure when tracking human activity. For the purposes of this thesis, Sec. Ch4.4 (within the article) and, to an extent, Sec. Ch4.3 are the most relevant and should be the primary focus, as they directly relate to the core topics discussed here. The remaining sections provide a broader, comprehensive review that, while valuable, is not immediately pertinent to the specific context of this document.

Publication details:

- Manivannan, A., Chin, W. C. B., Barrat, A., & Bouffanais, R. (2020). On the challenges and potential of using barometric sensors to track human activity. Sensors, 20(23), 6786.

Summary: This study presents a comprehensive exploration of the challenges and potential of using barometric sensors for human activity recognition, particularly focusing on vertical displacement activities. This work contributes to the growing field of Human Activity Recognition (HAR) by demonstrating how barometric pressure data, when combined with other sensors such as accelerometers, can significantly improve the accuracy of tracking vertical movements like stair climbing or elevator use. The paper is intended for researchers and practitioners in the fields of wearable technology, smart cities, and urban mobility, offering insights into how sensor fusion can enhance the precision of human movement tracking in both personal health and urban planning applications.

An essential contribution of this paper is the detailed investigation (through literature review and experiments) of the external factors that influence barometric pressure readings, particularly in real-world settings. These include variations due to weather conditions, changes in air velocity during movement, and the effects of built environments, such as transitions between indoor and outdoor spaces or the presence of HVAC systems. Understanding these factors is key to improving the reliability of barometers in recognizing vertical displacement activities. By mitigating the effects of external influences, this research enhances the use of barometric sensors for tracking human activity in diverse and dynamic environments, enabling more accurate applications in fields such as health monitoring, urban design, and mobility analysis.

sensors

MDPI

Review

On the Challenges and Potential of Using Barometric Sensors to Track Human Activity

Ajaykumar Manivannan [1], Wei Chien Benny Chin [1] and Alain Barrat [2,3,†] and Roland Bouffanais [1,*,†]

[1] Engineering Product Development, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore; manivannan_ajaykumar@mymail.sutd.edu.sg (A.M.); benny_chin@sutd.edu.sg (W.C.B.C.)
[2] CNRS, CPT, Aix Marseille University, Université de Toulon, 13009 Marseille, France; alain.barrat@cpt.univ-mrs.fr
[3] Tokyo Tech World Research Hub Initiative (WRHI), Tokyo Institute of Technology, Yokohama 226-8503, Japan
* Correspondence: bouffanais@sutd.edu.sg; Tel.: +65-6303-6667
† These authors contributed equally to this work.

Received: 17 September 2020; Accepted: 23 November 2020; Published: 27 November 2020

Abstract: Barometers are among the oldest engineered sensors. Historically, they have been primarily used either as environmental sensors to measure the atmospheric pressure for weather forecasts or as altimeters for aircrafts. With the advent of microelectromechanical system (MEMS)-based barometers and their systematic embedding in smartphones and wearable devices, a vast breadth of new applications for the use of barometers has emerged. For instance, it is now possible to use barometers in conjunction with other sensors to track and identify a wide range of human activity classes. However, the effectiveness of barometers in the growing field of human activity recognition critically hinges on our understanding of the numerous factors affecting the atmospheric pressure, as well as on the properties of the sensor itself—sensitivity, accuracy, variability, etc. This review article thoroughly details all these factors and presents a comprehensive report of the numerous studies dealing with one or more of these factors in the particular framework of human activity tracking and recognition. In addition, we specifically collected some experimental data to illustrate the effects of these factors, which we observed to be in good agreement with the findings in the literature. We conclude this review with some suggestions on some possible future uses of barometric sensors for the specific purpose of tracking human activities.

Keywords: barometer; barometric pressure; human activity recognition (HAR); vertical displacement activity (VDA)

1. Introduction

1.1. Development of Barometric Sensors

Barometers have been around for a very long time. While the air was thought to be weightless till the early 1640s, this changed when the Italian physicist and mathematician Evangelista Torricelli showed that a column of air exerts a significant force that can be measured by the amount of liquid displaced by the pressing air. This led to the discovery that "air has weight" and the invention of a measurement device that quantifies the atmospheric pressure [1]. In the latter part of 1640s, Blaise Pascal perfected the experiment and showed the finiteness of air pressure, leading to the hypothesis that the height of the atmosphere itself is finite, and to the proposition that altitude can

be measured as proportional to the atmospheric pressure [2]. The SI-derived unit, the Pascal (Pa), is named after Pascal's contributions to hydrodynamics, and is now officially used to measure the force applied by an atmospheric column of air above a unit surface area.

For the first two centuries after the invention of the barometer, this device was constructed using glass tubes filled with liquids, such as water or mercury [3]. In 1844, a new design appeared with the development of the aneroid barometer, which is purely mechanical, does not contain liquids, and shows the measurement value on a face dial [4]. These devices took a quantum leap with the advent of micro-fabrication in the 1960s, which allowed the miniaturization of the barometer and accelerometer to a size smaller than 0.1 mm—what is now commonly known as a microelectromechanical system (MEMS). With the successive advancements in integrated circuits and digitization of the sensor readings, the manufacturing and computational costs of these miniaturized devices were significantly driven down, thus paving the way for their widespread adoption in consumer products, especially in mobile phones. Currently, MEMS-based barometers are by far the most commonly found type of barometer in wearable devices and smartphones. In 2015, one of the pioneers in MEMS manufacturing, Bosch (Robert Bosch GmbH), claimed to manufacture one billion MEMS sensors per year for automotive and non-automotive applications in one production facility in Germany [5]. This company also claims to have their MEMS devices (including barometers) embedded in every second smartphone in the world [5]. Although this claim cannot be independently verified, it still points to the massive scale of production for this type of sensor, and underscores the ubiquitous availability of MEMS barometers.

Historically, barometers were used for weather forecasting and, thus, chiefly as environmental sensors. As a measurement device of ambient pressure, barometers have recently been used to measure evapotranspiration (transfer of water from land to atmosphere) in a given environment, for improving motor vehicle engine efficiency by modifying the air–fuel mixture, and for counting steps based on the slight disturbances in air pressure during body movements [6]. Thanks to the relation between pressure and altitude, barometers are also widely used as altimeters to measure altitude, particularly in airplanes. A number of recent applications are directed to the tracking of human activities, and there are still significant opportunities for the use of barometers in a vast range of additional applications. The full potential of barometers has not yet been taken advantage of, particularly in the Internet of Things (IoT) realm and with future consumer devices, particularly in the fast-growing area of wearable devices.

1.2. Barometers for the Tracking of Human Activity

In particular, the ubiquity of MEMS barometers in smartphones and other wearable devices makes them natural candidates as data sources for the study of human activities and for the field of human activity (and movement) recognition (HAR) [7]. Broadly speaking, HAR consists of using data from various types of sensors carried by individuals to automatically understand what type of activity they are carrying out. It consists, therefore, of choosing sensors that will be influenced by the activity, annotating a certain amount of data (that will serve as "training data") thanks to ground truth knowledge (i.e., the knowledge of the precise conditions in which the data were collected), and devising a classification task using typical machine learning frameworks to classify the rest of the data. The human activities considered can be broadly classified under two main classes: (1) ambulation and (2) transportation [7]. Ambulation refers to all movements and idle states of our human body (walking, idle, running, sitting, etc.), while transportation refers to our movement using vehicles (cars, buses, bicycles, etc.) [7]. In HAR research, such activities are considered under two conditions: (1) natural ones and (2) laboratory conditions. These conditions lie at the two extremes of a spectrum: "Natural" refers to individuals behaving normally within their usual environment, without any defined procedure and without being influenced by their being monitored, while laboratory conditions refer to a set-up that is especially designed for a human subject, who is given explicit instructions to perform a given activity. In reality, most experiments with human subjects happen somewhere in between these

two extreme conditions. It is also important to note that the commonly tracked activity classes vary for healthy and differently abled individuals. For example, tracking patients with movement disorders requires detection of distinct movement features, such as tremors, myoclonus, etc.

Until the 1990s, it seems that there is almost no reference in the literature to the use of barometers in tracking human activities, the most commonly used sensors being inertial measurement units (IMU)—comprising the accelerometer and gyroscope, along with the magnetometer. For instance, smartphones are equipped with accelerometers, whose signals are widely used to recognize most activity classes [8]. The situation changed with the silicon and digital revolution, which contributed to the effective use of barometers to track a range of human activities in the late 1990s. Initially, consumer devices such as mobile phones were equipped with barometers to improve GPS-based localization by reporting altitude or altitude changes [8]. Today, barometers along with a suite of sensors like IMUs and magnetometers are used—individually or through sensor fusion—to track a wide range of human activities. Such use of a barometer to track human activities is a fairly recent phenomenon. In the last two decades, barometers were found to improve some activity class recognition that involves change in height, such as falls or vertical movements. In some applications, such as recognizing vertical displacement activity (VDA), the accelerometer has been replaced by or at least given less importance than barometers, which are more energy efficient, require less signal processing, and yield less noisy signals than IMU signals. Most smartphones have barometers allowing them to predict changes in altitude with an accuracy of the order of one meter. However, an effective tracking of human activities is best obtained with the combined use of these important sensors, each providing unique information on the subject's state.

It is, however, important to note that our ability to properly leverage the potential of barometers for HAR purposes critically hinges on our understanding of the physical properties of the atmospheric pressure. Indeed, the measure of the ambient pressure by a barometer is influenced by the static and dynamic properties of its environment [9]. Its effective use as a signal thus requires dedicated data post-processing techniques and classifiers in general to account for external factors: For instance, if one is tracking altitude changes and vertical displacement activity, the variations in the local atmospheric pressure have to be accounted for. Measures are moreover affected by the sensor itself, whose accuracy and manufacturing imperfections can introduce noise and variability between devices. Hence, it is fair to say that the numerous factors affecting barometric pressure (see Section 4), if not properly understood and accounted for, can hinder the effective use of barometers to detect and identify particular classes of human activity.

In order to systematically discuss the issues related to the use of barometric pressure to track human activities, we have collected and organized a series of scientific articles to understand the potential and limitations of such an endeavor. The current review is non-exhaustive but intentionally limited to the use of barometric sensors for the most common classes of human activity. This choice is justified by the wide range of applications offered and the fact that barometers now pervade many mobile devices (wearables and smartphones, cars, etc). The applications explored are also primarily limited to the recent developments in using MEMS-based barometers for consumer goods/electronics. It is worth stressing that this review does not address the vast breadth of machine learning (ML) or other advanced classifiers used to interpret the sensed data or to identify a given activity with a given accuracy. However, the details provided in this review are most useful for the further development and design of effective ML/classifier strategies.

This review is organized as follows: Section 2 introduces the general sensor data collection process to track human activity and discusses a number of general issues common to many types of sensors. Section 3 focuses more specifically on the use of barometers in human activity and movement recognition. Section 4 describes the factors that affect barometric pressure and quantifies the order of magnitude of each effect based on a range of studies reported in the literature. This section is enriched by data especially collected for illustration purposes. Lastly, Section 5 reviews the key findings of

the above sections, explores the current challenges, and recommends future directions for different applications of barometric sensing applied to tracking a range of human activities.

2. Tracking Human Activities: Background about Data Collection

Three components are key to studying and understanding human activities, namely (i) motion tracking, (ii) recognizing the type of activity, and (iii) analyzing the obtained patterns. Barometers are part of the suite of sensors used in each of these steps. As such, the characteristics of the data collection process are similar to those for other types of sensors. They include the characteristics of sensors, their placement and orientation, the sampling frequency, and the environmental conditions. The type of application (activity class, diversity of sample population) and the method for recording the ground truth and annotating a part of the data also have to be carefully designed.

Data collection and annotation are indeed critical to the effectiveness of subsequent stages of classification—e.g., data pre-processing, feature engineering and identification—in the overall workflow of sensed data associated with human activity recognition. Similarly, an informed decision on the data collection procedure depends on the specific problem under study and on the particular class of human activity being investigated. Depending on the sensor characteristics and the class to be recognized, data collection methods must be tailored to shed light on the phenomena under investigation for better accuracy and performance. The success of such data-processing activity not only depends on acquiring the data, but also on being able to effectively process it and extract meaningful features and patterns. In this section, we successively discuss the key elements of data collection to track human activities: the commonly used sensors and sensor suites, the placements and orientation of sensors and their effects, the sampling frequency for tracking and detecting the activity classes of interest, the types of activity classes related to human movements, the nature of data in terms of the levels of realism (natural, laboratory, and in between), the class imbalance issues, the diversity in sampling targets (users), and, lastly, the challenges in annotation techniques.

2.1. Sensors and Sensor Suites

Smartphones and watches are by far the most common and straightforward way to collect data in a natural setting. Nonetheless, some research groups studying particular behaviors and phenomena required a custom-built wearable sensor design for both laboratory testing as well as for operations in real-life conditions. For instance, three-axis accelerometers are the most widely used sensors [10,11], followed by three-axis gyroscopes and magnetometers. Since these inertial sensors are now commonly found in today's smartphones and other MEMS devices, many recent studies use a combination of these sensors to improve classification accuracy [12,13], although a three-axis accelerometer alone can extract good-quality data, resulting in excellent classification results [14]. Wearable devices may also contain environmental sensors that measure, for instance, temperature, light, atmospheric pressure, and sound to assist in context detection, and/or physiological sensors, such as for the heart rate for medical research [7] and personal fitness purposes. In transport mode recognition, GPS is the most widely used sensor [15], followed by telecommunication data [16,17], WiFi access points [18,19], and travel surveys [20]. In general, location-based sensors operate based on a combination of inertial sensors to build robust recognition systems that can distinguish between different travel modes, including walking and being idle [21]. Fusion of environmental sensor data with inertial and location-based sensed data provides sufficient information to detect what is commonly known as 'Activities of Daily Living' (ADL). In this context, barometric pressure sensors have traditionally been considered as environmental sensors used to measure ambient pressure. However, they are also capable of sensing movement/activity in ways similar to inertial sensors, especially when considering vertical movements.

2.2. Placement and Orientation of Sensors

The placement and orientation of the sensor might influence the characteristics of the captured signal, thereby affecting the recognition accuracy: Indeed, the training data might then fail to account for all the possible variations, often resulting in a sparse feature space. Numerous studies have focused on ways to alleviate this particular effect, and have provided solutions that range from collecting diverse data to independent features not affected by such parameters [22–24]. Chen et al. [25] proposed the use of coordinate transformation along with principal component analysis (PCA) to reduce this issue associated with orientation changes. Several groups have also studied the question of optimal sensor placement [26,27]. Interestingly, while this issue affects many sensors, this is not the case for barometric pressure sensors, whose readings are widely independent of their on-body position and orientation [28], even if they are dependent on a range of environmental conditions.

2.3. Sampling Frequency for Capturing Human Activities

The temporal resolution of the data is directly related to the sampling frequency of the sensor used. For the vast majority of sensors used to carry out human activity recognition, the sampling frequency typically ranges from 10 to 100 Hz, with the rate going as high as 512 Hz. It is commonly reported that the characteristic frequencies of most human activities are below 10 Hz, and therefore, the optimal sampling rate—based on the Shannon–Nyquist theorem—is 20 Hz [29]. Khan et al. [30] reviewed five public datasets and showed that the sampling rate considered could be reduced between 48% and 86%, with a minimal sampling rate of approximately 12 Hz [30]. Yan et al. [31] also studied the effect of sampling rate on energy consumption, and they concluded that a higher sampling rate increases energy load without providing additional meaningful information, in agreement with the results reported in [30]. From the energy perspective, Yan et al. [31] concluded that a shorter dataset with a higher sampling rate is preferable to a longer dataset at a lower sampling rate. It is worth adding that the optimal sampling rate also depends on the type of human activity to be recognized [30] and on the type of sensor used. For instance, when studying large-scale human movement, sampling GPS data or other location signals at rates comparable to those of accelerometers is unnecessary. Hence, the GPS signals are typically sampled at around 1 Hz [32,33]. For pressure signals, barometers have been sampled at rates as low as 1 Hz [34].

2.4. Classes of Human Activity

Lara and Labrador [7] thoroughly reviewed a comprehensive list of human activities recognized and categorized in the literature, including ambulation, transportation, gestures, exercises, and daily living activities. Ambulation refers to all states of movements performed directly by a human, e.g., walking, idle, running, and sitting. Transportation refers instead to one's movement when riding vehicles, including cars, buses, and bicycles. Gestures also correspond to body movements, such as picking up or putting down objects, looking at watches, moving the head, etc. Exercises comprise the motion of the human body for the purpose of physical health and activity (jogging, running, playing ball, etc.). Daily living activities are a group of activities that most people experience on a daily basis (e.g. eating, drinking, and reading). The activity classes may, of course, occur concomitantly and may, therefore, be composite [35], interleaved, concurrent, or overlapping [36]. The design of data collection—e.g., the selections of sensors, sampling rates, and recognition approaches—is heavily dependent on the types of activity classes to be recognized. For data with low sampling frequency, i.e., lower than 1 Hz, more than one activity could be performed during the same time interval. In particular, the class of vertical displacement activities (VDA) has generally been recognized in the literature as part of the larger class of ambulation activities. Similarly to accelerometers, barometers are very well suited for the recognition of VDA. In addition, barometers can accurately determine altitude changes for VDA occurring at a sampling rate lower than 1 Hz.

2.5. Nature of Data

The data used to study human activity can be broadly grouped under two extreme conditions: (1) natural and (2) laboratory. While the former concerns data collected during real-life activities of people in a given city or country, and might thus be affected by complex environmental factors and all kinds of unknown conditions, the latter corresponds to activities performed in fully controlled environments with known conditions. Several groups have collected their own datasets to fit the specific needs and requirements of their studies [37,38]. Many benchmark datasets are also publicly available, and are commonly used to validate new methods [39–42]. The recent review article [43] reported details of several key benchmark public datasets and provided a rich analysis on the content and application-context studies. The datasets collected and made publicly available were either collected in laboratory conditions [44] or are real-world ones—where participants had instructions to perform specific activities given implicitly [45] or explicitly [24], or were instead allowed to move freely while being unobtrusively observed [13]. These datasets have been claimed to be semi-naturalistic or realistic even when the subjects were given scripted instructions [46,47]. Lara and Labrador [48] reported on subjects performing interleaving activities as they naturally occur, in sequence rather than segmented, to reach realistic conditions. Truly realistic data correspond to data collected while people go about their activities of daily living (ADL) [14]. As highlighted in [49], the use of purely realistic data would, however, require a long data collection period. Hence, segmented data obtained by a set of experiments corresponding to various activity classes are typically collected for training. Some studies such as [50,51] have managed to deal with this issue by considering segmented data for training and activities of daily living or sequence data for validation.

2.6. The Class Imbalance Issues in the Tracking of Human Activity

The problem of class imbalance in datasets—the fact that some activities are more prevalent than others and are, therefore, over-represented—is a recurrent condition found in both public and private datasets and known to affect recognition accuracy [47,52,53]. For instance, data corresponding to physically tiring activities, or that are difficult to obtain for other reasons, naturally form an under-represented portion of the data, thereby leading to inherent data imbalance [38]. This issue can be offset by oversampling the minority class or under-sampling the majority class. Chen and Shen [38] used the so-called synthetic minority over-sampling technique (SMOTE) for oversampling the minority class, while Nguyen [54] used a modified version of it. Guan et al. [52] also proposed the use of an ensemble of deep learning models to offset such a data imbalance. Other studies have tackled the problem by simply using F1-score to report classification accuracy that takes into factor the different sizes of each class [52,55]. This issue can occur in applications using barometers, as the number of altitude changes in human movement for any given continuous experiment is usually very limited, and these changes typically have, moreover, a rather high inter-event time in ADL.

2.7. The Physical Characteristics of Carriers

The physical characteristics of the sampled users/carriers directly affect the measured human activity patterns; thus, they are an important aspect of the data collection process for HAR. Data collection methods can contribute to the generalization in recognition models by allowing diverse characteristics or parameters to be incorporated into the training dataset. These characteristics should be representative of the final dataset to which the model is applied. One of the important parameters is the user themself, whose diverse physical characteristics can be challenging to integrate or model [37]. Studies generally report the user's age, height, weight, body mass index (BMI), physical ailments, etc. for context and applicability of the study [55–58]. In the case of barometers, the height of the user/carrier has only a limited effect on the use of a barometer, and the resolutions of most wearable sensors are within the range of placement difference. It is thus safe to assume that the barometer sensor works independently of the physical characteristics of the carrier.

2.8. Annotation Techniques

Stikic et al. [59] note that the accuracy of annotation is subject to a trade-off between length of the data collection and the time and effort required for labeling the dataset. Indeed, if direct observation is required for accuracy, this results in a prohibitively expensive requirement. De la Hoz Franco et al. [43] carried out a meta-review of 374 papers on human activity recognition (HAR) and found that 60% of the used data were annotated. However, most studies record ground truth by resorting to experiments in a laboratory setup with the help of researchers and supporting infrastructure [27,57,58]. Some studies have also relied on subject self-reporting [37,60,61], which is known to be error-prone due to the obvious difficulty in marking precise times while carrying out the activities of daily living. Chung et al. [13] recorded ground truth by unobtrusively following the subjects one at a time. Video capturing [34,47,55,62], audio recording [63], and GPS localization have been highlighted as methods of direct observation without requiring one to monitor the subjects under constrained and controlled conditions. Recording ground-truth data for long-term studies of ADL has been repeatedly acknowledged to be impractical and/or prohibitively manpower intensive [49,64]. For instance, Willetts et al. [14] used automatic cameras to record ground truth every 20 s for 143 participants over 24 h. This shows the interest of automated methods, such as the active learning methods developed by Bota et al. [50], which require the manual annotation of a small subset of the data, while the rest is automatically labeled.

A practical, privacy-sensitive, and unobtrusive annotation method for studies using barometers to track humans over long durations is manual labeling of the data. Unlike the signal sensed by accelerometers or other inertial sensors, barometric pressure can indeed be less complex to interpret. However, this process comes with several challenges, one of which is the lack of complete understanding of all the factors that influence barometric pressure. This important issue is addressed in detail in Section 4.

3. On the Use of the Barometer in Human Activity Recognition

For the sake of studying human activity recognition, the barometer has been primarily used to measure changes in altitude (or elevation). The scale of the altitude changes varies from a fall to vertical displacements like moving uphill, climbing a deck of stairs, riding an elevator, etc. In recent applications, the patterns of the time series of barometric signals have been shown to be a good indicator of the underlying activities, such as walking, being idle, and using transportation. The rate of pressure change can help to identify the mode of vertical transport and determine the vertical velocity of air vehicles. Barometers are widely embedded in wearable devices and used for vertical transport detection [28,65–70], indoor positioning and navigation [68,71–73], building monitoring [74], health monitoring [75,76], vehicle tracking [77,78], transport mode detection [79], and GPS localization improvement [68].

One of the earliest works on using barometers to classify vertical movement is due to Sagawa et al. [80] and dates back to 1998. In [80], Sagawa et al. collected 83 min of data using both an accelerometer and barometer at a 100 Hz sampling frequency from six males between 20 and 40 years of age. Their classification model was trained offline using cut-off values selected heuristically. Since then, recognition and identification procedures have greatly improved thanks to a number of factors, including new sensor types, larger range of sampling rates, user and device characteristics, modes of carriage, power consumption, real-time demand, classification models, and, finally, ground truth availability.

3.1. Barometric Pressure Sensor for Tracking Human Activity

MEMS barometers are miniature sensors (<0.1 mm) manufactured by prominent companies, such as Bosch [77,81,82], ST Microelectronics [77,81,83], and Measurement Specialities [84]. As mentioned previously, they are commonly found in wearable devices and smartphones. They have

high precision but relatively lower absolute accuracy than table-top barometers. Specifically, the accuracy in measuring absolute pressure is low in these sensors, but their relative pressure accuracy is shown to be as low as $\pm$1.2 Pa [84,85], with average mobile devices having a resolution of $\pm$12 Pa [28]. It is worth adding that the measured absolute pressure varies from one device to another due to manufacturing differences and other technical factors [81].

3.2. Data Processing for Sensed Barometric Pressure

Barometers are constrained by their resolution and sampling frequency, which limit precision and accuracy, and the use of their data is further constrained by the presence of noise. Some studies have used filtering and signal modeling to overcome the latter problem. For instance, moving average filters over a given time window is widely used for that purpose [71,84,86], followed by other finite impulse response (FIR) filters [87] and infinite impulse response (IIR) filters [76,87–89], such as double exponential smoothing [28,69]. Signal modeling, like the sinusoidal fitting model [90] and sigmoidal nonlinear fitting, is commonly used to increase the contrast of elevation changes [91].

The time series data of barometric pressure are generally converted into statistical [70,92], spectral [70,92], temporal [70,92], or wavelet-based features [75,93] before being fed into a classifier. These features are designed to enhance the detection of the specific activity of interest. Most of the features used are based on the rate of change of pressure [69,70,79,84,94] (also known as vertical velocity, or simply the slope) and the standard deviation of differential pressure (dp) [74,79], which differentiates altitude changes from other environmental factors that influence ambient pressure.

3.3. Classifiers for HAR

The choice of the classifier depends on the application at hand. It also depends on the range of pressure variations and durations under consideration. Some studies have only considered indoor pressure profiles of individuals [28,68], while few have considered the full range of activities of daily living (ADL), including both indoor and outdoor events like transportation [77,79]. A person carrying a barometer is indeed affected by factors present outside their specific environment, which seriously limits the scope of many studies of HAR using barometer data.

Barometric pressure is more straightforward than inertial sensors in conveying sensed information due to its fairly direct reading, which greatly simplifies the use of classifiers. The most widely used classifiers are decision trees [28,65,68,82,83,87], support vector machines (SVMs) [75,77,95,96], and threshold-based models [76,91,97,98]. Clustering models, such as hierarchical clustering [73,99] and k-Means clustering [71], Bayesian-based classifiers [72,100,101], LSTM models [28,102], and fuzzy inference models [90] have also been used. Hidden Markov models take advantage of logical activity sequences that can be associated with some activity sequences (such as riding an elevator before and after walking) [103], while fuzzy inference models take advantage of context and behavioral constraints [90].

Liu et al. [69] compared classifiers such as Random Forest, J48 decision trees, artificial neural networks (ANNs), SVMs, and Naïve Bayes to classify horizontal displacement activity from vertical displacement activity; the Random Forest classifier was found to have the highest accuracy. Vanini et al. [28] compared the performance of Bayesian networks, decision trees, and recurrent neural network (RNN) models to recognize VDA. It has been found that RNNs have a 99% accuracy, while decision trees provide the optimal trade-off in terms of computational cost, energy efficiency, and accuracy.

Some applications do not require any classifiers. For instance, state estimation, like altitude and vertical velocity, can be determined using a Kálmán filter or one of its many variants [80,86,96,104,105]. As another example, Bollmeyer et al. [84] solely used the differential pressure to estimate altitude. Ho et al. [78] used the so-called dynamic time warping (DTW) technique to establish the correlations between the pressure time series data and known geographical elevations to track vehicles. Similarly, Hyuga et al. [106] found ways to use the variance of differential pressure to account for the variations

in barometric pressure associated with air velocity and the built environment; they used similarity measures between pressure and known altitude to locate a train/user in a rail network.

3.4. Applications: Barometer-Only Studies

Very few studies dedicated to tracking human activity and/or movement solely rely on the use of a barometer, despite the fact that their potential for applications is promising. The selected applications of using only a barometer for HAR are summarized in Table 1, in which we highlight the role of barometers in each study, the considered factors that may affect the performance of barometers, and the key findings of the studies. In addition, the types of human activity (i.e., the activity classes: (A) ambulation and/or (T) transportation), the built environment (i.e., locations: indoor and/or outdoor), and the time period of activities (short term and/or long term) were categorized for comparison purposes. These applications were selected based on three criteria: (1) They use only barometers as sensors, (2) they are representative of their field/sub-field or add unique methodology or results in their field, and (3) they add to the understanding of the many factors that influence barometric pressure. Most of the studies focused on ambulation activity (10 out of 16), while two of them focused on transportation activity, two of them studied both, and the rest (2) were neither about ambulation nor transportation activities. Among the 16 studies, nine were performed in an indoor environment, three in the outdoors, and the other four considered both indoor and outdoor environments and the corresponding factors. Most applications (11 out of 16) focused on the changes in barometers' readings in a short-term period, two studies focused on long-term patterns, and the other three considered both short- and long-term evolution. Based on these 16 studies, the main factors that may affect the performance of barometers included the altitude of the activity, the background climate and weather, the surrounding built environment, the relative air velocity due to motion, and the sensor accuracy. We discuss the results of these studies in more detail in the following paragraphs.

Barometers have been used to measure altitude as a stand-alone measuring instrument for a very long time [81]. The surface of the land on which we move and travel, including transportation routes, is uneven. This topographical feature can be estimated using barometric pressure, and the subtle changes in elevation along the travel routes can be exploited for localization of people riding vehicles. As stated before, this is, however, challenging due to two major factors that affect barometric pressure—built environment, like tunnels, bridges, etc., and the air velocity during motion. However, Hyuga et al. [106] used the pauses of trains in stations between successive train rides to locate a user/train in a subway route by computing the successive altitude changes and comparing them with known relative elevations of train stations.

Similarly to accelerometer data, the signal pattern encoded in a barometer output carries sufficient information to recognize a range of human activities. Ghimire et al. [87] observed the change in air pressure when a person walks with hands swinging and used this gait pattern to count steps. The gait pattern is also used to detect the walking class with approximately 95% accuracy [87], similar to the performance of available accelerometer-based recognition methods [63,107], and this accuracy fluctuates for both sensors based on the on-body sensor position. In this application, the barometric data will also be person-dependent like with an accelerometer, since the gait patterns detected from air pressure changes can vary among the population. The pressure fluctuations due to vehicle motion were used by Sankaran et al. [79] to differentiate between the distinct patterns produced by people riding vehicles as opposed to walking or standing idly.

In climate-controlled buildings, changes in barometric pressure can be detected during indoor-to-outdoor transitions. Wu et al. [74], moreover, showed the possibility to detect the opening or closing of a building's entrance doors, even with a barometer located far from the doors, and even analyzed the patterns to determine the type of door (automatic or manual). They highlighted implications for building monitoring and security, along with the potential application of tracking human movement in a building.

Table 1. Categorized literature related to the application or use of barometers for human activity recognition. Two types of activity classes: (A) ambulation and (T) transportation.

Ref.	Use of Barometer	Factors Considered	Contributions and Applications	Activity Class(es)	Location	Time Period
[81]	Estimate altitude and altitude changes	Altitude, climate and weather, and sensor accuracy	Evaluate sensors to estimate the altitude of airplane above ground and the orientation angle of wings using dual-device systems	–	Indoor	Short
[106]	Estimate altitude and altitude changes	Altitude, air velocity during motion, and built environment	Estimate the location of a traveler in a subway using only a barometer	T	Outdoor	Short
[87]	Detect gait patterns and estimate altitude changes	Altitude	Step detection and activity recognition including VDA using a barometer	A	Indoor	Short
[79]	Detect vehicle patterns and altitude changes	Altitude, climate and weather, air velocity during motion, and sensor accuracy	Identify transportation modes and ambulation activities using a barometer	A & T	Indoor and outdoor	Short
[74]	Detect door opening/closing in building and estimate altitude changes	Altitude, climate and weather, and built environment	Detect door opening/closing to monitor building activities and recognize VDA	A	Indoor and outdoor	Short & long
[108]	Estimate altitude and altitude changes	Altitude, climate and weather, built environment, air velocity during motion, and measurement accuracy	Recommendations to build indoor localization from reference pressure	A	Indoor	Long
[66]	Estimate altitude and altitude changes	Altitude, climate and weather, built environment, and sensor accuracy	Floor localization using reference pressure from multiple barometers in each floor	A	Indoor	Short
[99]	Estimate altitude and altitude changes	Altitude, climate and weather, and sensor accuracy	Calibration of wearable barometers using crowd-sourcing to enable floor localization. No knowledge of building or additional infrastructure is required	A	Indoor	Long

Table 1. *Cont.*

Ref.	Use of Barometer	Factors Considered	Contributions and Applications	Activity Class(es)	Location	Time Period
[84]	Estimate altitude and altitude changes	Altitude, built environment, and sensor accuracy	Studied the different factors that affect barometric pressure in the built environment. Estimate indoor altitude	A	Indoor	Short
[68]	Estimate altitude changes and mode of vertical transportation	Altitude, climate and weather, built environment, and sensor accuracy	Identify VDA and mode of vertical transport	A	Indoor	Short and long
[104]	Estimate altitude	Altitude, climate and weather, and sensor accuracy	Estimation of altitude for indoors and outdoors	A	Indoor and outdoor	Short
[28]	Estimate altitude changes	Altitude, climate and weather, built environment, and sensor accuracy	Activity recognition including VDA using only barometer and comparison with accelerometer-only and GPS-only approaches	A & T	Indoor and outdoor	Short
[85]	Estimate altitude	Altitude, climate and weather, and sensor accuracy	Barometer measurement error modeling and correction to track air vehicle	–	Outdoor	Short
[78]	Estimate altitude and altitude changes	Altitude, climate and weather, built environment, air velocity due to motion, and sensor accuracy	Compared barometric pressure data with topographical elevation data to localize and track vehicles	T	Outdoor	Short and long

Applications that use only barometer data are more common in the studies concerned with floor localization and recognition of VDA. In the absence of location sensors, the challenge with floor localization is to have a reference pressure and associate it with the data measured by the considered wearable devices at the moment they enter a given building. To obtain such a reference pressure, Li [108] recommends receiving it from a location that is similar to the environment in which it is deployed; this setup is important, as the reference pressure obtained from reference stations can potentially experience different environmental effects. Xia et al. [66] showed that the barometric pressure pattern can change from floor to floor in idle settings, and hence, installed a calibrated barometer in each floor to collect multiple reference pressures. Ye et al. [99] applied an infrastructure-independent approach by constructing an encounter network—determined by comparing simultaneous pressure changes—and use a root node to calibrate all the mobile sensors. This method is, of course, prone to errors, and the lack of calibration has been shown to result in an accuracy of only 70%.

The challenge for recognizing VDA using only barometer data is that the sensor data should ideally be free from all factors that affect barometric pressure other than altitude. This can be guaranteed if the sensor data are collected from a controlled environment where no other factors that can be mistaken for VDA occur. As shown by Bollemeyer et al. [84], Ghimire et al. [87], and Muralidharan et al. [68], this can be achieved by limiting the studies to indoors, where only weather, sensor accuracy, and built environment effects impact the barometric readings. Liu et al. [104] collected experimental data from outdoors, like mountain climbing, while avoiding activities that cause adverse pressure gradients, like transportation. Vanini et al. [28] demonstrated their VDA recognition capability while considering ambulation and transportation, with, however, a limited set of activities, including, in particular, only cable-cars as an outdoor transportation mode.

In summary, in a specific environment and considering selected modes of activity, the recognition performance of VDA using barometer can be very high. More generally, in any given environment, barometers are shown to perform far better or similarly to other sensors when detecting VDA, and represent the only viable way to extract the magnitude of vertical displacement. Vanini et al. [28] report that with barometer data, one achieves similar performance in classifying VDA compared to with an accelerometer or GPS, but a superior one in energy efficiency and independence in terms of sensor location and orientation. Muralidharan et al. [68] similarly showed that barometer-only classification is significantly more accurate (99%) in recognizing modes of VDA compared to accelerometer-based classification (85%). However, the accuracy for accelerometer-based classification drops below 30% when the mobile phone is used for taking calls or playing games.

3.5. Applications: Multi-Sensor Studies

Few studies have employed a barometer as the sole sensor in their application. It is usually integrated with other sensors, like inertial sensors [69,82], environmental sensors (light, temperature, sound, etc.) [109,110], location-based sensors (GPS) [71,96], and communication infrastructure (WiFi, Bluetooth, RFID, etc.) [72,73,100,110]. Several studies are dedicated to improving the sensor fusion of inertial sensors with barometers, which constitutes a critical step in the optimization of activity recognition [86,110,111].

To summarize and categorize the vast breadth of applications of barometers for HAR with multi-sensors, we have gathered some selected studies into Table 2. The studies were selected based on the same criteria as in Section 3.4. Table 2 contains a list of (29) studies that include barometers as part of their multi-sensor set-up. We detail in the table which other sensors are used, the particular role of barometers in the studies, the factors that may affect the barometric readings, and the key contributions of the studies. Similarly to Table 1, we also categorized the studies by activity classes (A: ambulation, T: transportation), built environment (location, including indoors and/or outdoors), and periods of the experiment (time period, including short term and/or long term) for comparison purposes. Most of the studies used barometers together with accelerometers (23 out of 29, 10 of which

were using only these two types of sensors), gyroscopes (12 out of 29), and magnetometers (7 out of 29). Other sensors, such as WiFi fingerprinting (six studies), GPS (two studies), and Bluetooth (one study), were used for indoor or outdoor positioning; in addition, light and foot pressure—both were used in one of the 29 studies—were also included for some specific aims. Most of the applications focused on ambulation activities (26 applications), one of the 26 included cycling activity, and another one included transportation activity; only one of the other three studies focused solely on transportation, and the other two could not be categorized. Almost all of the listed applications were conducted in an indoor environment (27 studies), six of which were also tested in an outdoor environment; one study was conducted only outdoors, and in another one, the context was not mentioned. In terms of duration, most studies focused on short-term activity (28 studies), three of which studied both short-term and long-term human activities; only one application focused solely on long-term activity.

In most classical HAR analyses, the classification accuracy in detecting VDA based on sensory data without pressure is usually low [89]; this is not seen as a critical issue, since VDA is not the focus or priority. Increasingly, the barometer has been recognized as an important sensor in HAR, where accurate recognition of VDA is critical to many applications [105]. Hence, a majority of applications that aim to measure altitude or track altitude changes employ a barometer as part of their sensory suite. For instance, in health monitoring applications, the inclusion of the barometer has been shown to improve VDA recognition [44,75,95], fall detection [76,83,93], and estimation of energy expenditure and physical activity [44,91,97,112].

Accelerometers are still the predominantly used sensors in HAR, and have been widely used as stand-alone sensors in recognizing many activities of daily living. They complement barometer-based recognition algorithms in detecting ambulatory movements such as walking, and their use also helps distinguish stair-climbing from other modes of vertical transportation, like using an elevator and escalator [69,101]—and even elevator from escalator [67]. Sankaran et al. [79], however, point to the high cost associated with the use of accelerometers: demands in data acquisition (position and orientation dependent), high sampling rate, complex processing, and classification training.

For indoor localization and navigation applications, obtaining a reference pressure to calibrate all mobile sensors is critical for the system to work. This is more easily obtained in multi-sensor applications. Pipelidis et al. [71] used light sensors to detect the transition between indoors and outdoors so as to derive a reference pressure at ground level, which subsequently serves the detection of floor levels. Communication infrastructures like WiFi, Bluetooth, and RFID are also used to provide additional location information to help assist indoor localization or transmit location-specific data, such as reference pressures in a floor, to assist barometers for calibration purposes. Tachikawa et al. [110] even combined WiFi signal with a microphone and other inertial sensory data to detect the type of indoor location—restroom, desk, elevator, etc.

Barometers are known to speed up the GPS localization through their altitude estimation [68]. Conversely, the altitude information can be accessed from GPS localization [104]. Furthermore, GPS or any location information can help distinguish the transportation modes from ambulation, where the changes in barometric pressure can easily be misunderstood for altitude changes. Even though the elevation changes are present in our transportation paths, the altitude estimation from barometric pressure due to air velocity during motion can be predominant. Some studies have used the barometric pressure instead of GPS to track people riding vehicles [44,77], but this can be very misleading, as the significant changes in pressure due to vehicle motion and the built environment like tunnels and bridges have not been fully taken into account or even properly understood.

Table 2. Categorized literature related to the application or use of multi-sensors with barometers for human activity recognition. Two types of activity classes: (A) ambulation and (T) transportation.

Ref.	Additional Sensors	Use of Barometer	Factors Considered	Contributions and Applications	Activity Class(es)	Location	Time Period
[69]	Accelerometer, Magnetometer, Gyroscope	Estimate altitude changes	Altitude, climate and weather, built environment, and sensor accuracy	Improved recognition of VDA using barometer	A	Indoor	Short
[82]	Accelerometer, Gyroscope	Estimate altitude changes	Altitude	Identify ambulation activities including VDA	A	Indoor	Short
[110]	Accelerometer, Gyroscope, Magnetometer, WiFi, Microphone	Estimate altitude changes	Altitude	Determine location semantics, such as restroom, desk, elevator, etc., using sensor fusion	A	Indoor and outdoor	Short
[71]	Light, GPS	Estimate altitude and floor level	Altitude	Vertical indoor mapping	A	Indoor and outdoor	Short
[72]	WiFi	Estimate altitude and altitude changes	Altitude, climate and weather, and sensor accuracy	Improved barometer measurement error model and sensor fusion for floor localization	A	Indoor	Short and long
[73]	WiFi	Estimate altitude	Altitude and sensor accuracy	Improved floor localization from crowd sourcing using few devices equipped with a barometer	–	Indoor	Short
[100]	WiFi	Estimate altitude and altitude changes	Altitude, climate and weather, and sensor accuracy	Floor level identification by hybrid approach between barometer-only and WiFi-only methods. The barometer-only approach uses crowd-sensed barometer data for self-calibration and builds an elevation map independently in each device	A	Indoor	Short
[111]	Accelerometer, Gyroscope	Estimate altitude and correct accelerometer errors	Altitude	Improved IMU–barometer sensor fusion	–	Indoor	Short
[86]	Accelerometer, Gyroscope	Estimate altitude changes	Altitude, climate, and weather	Improved sensor fusion to track vertical motions	A	Indoor	Short

Table 2. *Cont.*

Ref.	Additional Sensors	Use of Barometer	Factors Considered	Contributions and Applications	Activity Class(es)	Location	Time Period
[89]	Accelerometer, Gyroscope	Estimate altitude changes	Altitude	Improved accuracy in recognizing ambulation activities including VDA	A	Indoor	Short
[44]	Accelerometer	Estimate altitude changes	Altitude	Using a barometer improved overall activity recognition including VDA and better estimated of energy expenditure	A	Indoor	Short
[95]	Accelerometer	Estimate altitude and altitude changes	Altitude	Child activity recognition including VDA to prevent injuries	A	Indoor	Short
[75]	Accelerometer, Magnetometer, Gyroscope		Altitude, built environment, and sensor accuracy	Activity recognition including VDA for health monitoring of stroke patients	A	Indoor and outdoor	Short
[93]	Accelerometer	Estimate altitude changes	Altitude	Improved accuracy by including a barometer for fall detection	A	Indoor	Short
[76]	Accelerometer	Estimate altitude changes	Altitude	Improved fall detection using a barometer	A and T	Indoor	Short
[83]	Accelerometer	Estimate altitude changes	Altitude	Low-power fall detection for long-term monitoring	A	Indoor and outdoor	Short and long
[91]	Accelerometer, Gyroscope, Magnetometer, Foot pressure	Estimate altitude change	Altitude	Activity recognition including VDA for better estimation of elderly physical activity	A	Indoor	Short
[97]	Accelerometer	Estimate altitude changes	Altitude and sensor accuracy	Detection of VDA improved the estimation of physical activity and energy expenditure	A + cycling	Indoor and outdoor	Short
[112]	Accelerometer	Estimate altitude changes	Altitude, climate and weather, and sensor accuracy	Detection of VDA improved the estimation of energy consumption and physical activity	A	Indoor and outdoor	Short and long
[101]	Accelerometer	Calculate vertical displacement and estimate floor level	Altitude, climate and weather, and built environment	Floor localization	A	Indoor	Short
[67]	Accelerometer, Gyroscope	Estimate altitude changes	Altitude	Improved identification of VDA for indoor localization	A	Indoor	Short
[77]	GPS	Detect vehicle patterns	Altitude, climate and weather, built environment, and air velocity during motion	Use the effect of elevation changes in roads and air velocity due to motion to detect the motion state of a vehicle and help identify traffic congestion	T	Outdoor	Long

Table 2. *Cont.*

Ref.	Additional Sensors	Use of Barometer	Factors Considered	Contributions and Applications	Activity Class(es)	Location	Time Period
[92]	Accelerometer	Estimate altitude changes	Altitude	Reduced position and orientation dependency using a barometer	A	–	Short
[80]	Accelerometer	Estimate altitude changes	Altitude, climate and weather, and sensor accuracy	Earliest known activity classification including VDA using a barometer	A	Indoor	Short
[65]	Accelerometer, Magnetometer	Estimate altitude changes	Altitude, built environment, and sensor accuracy	Identify mode of vertical transportation for indoor navigation	A	Indoor	Short
[90]	Accelerometer, Gyroscope	Estimate altitude changes	Altitude and climate and weather	Improved activity recognition including VDA using a barometer by considering behavioral rules and applying context in a two-step process	A	Indoor	Short
[113]	WiFi, Bluetooth	Estimate altitude and altitude changes	Altitude, climate and weather, and sensor accuracy	Used WiFi/Bluetooth beacon to localize the user and recorded the respective pressure as the reference pressure for that floor. Any change in elevation is then used to identify the floor level	A	Indoor	Short
[114]	Accelerometer, Gyroscope, Magnetometer, WiFi	Estimate altitude and altitude changes	Altitude	3D indoor localization	A	Indoor	Short
[70]	Accelerometer, Gyroscope, Magnetometer	Estimate altitude changes	Altitude	Identify the mode of vertical transport	A	Indoor	Short

4. Factors Affecting Barometric Pressure

Atmospheric pressure is the force applied per unit area by a column of air above a specified area [115]. It is caused by the gravitational pull of air molecules to the Earth's surface. Atmosphericpressure measured by a barometer is also called the barometric pressure. The fact that the atmospheric pressure varies monotonically with altitude has made it a useful measurement tool for indicating heights, and the development of portable barometers has extended the usage of barometric pressure to the indication of the carrier's position in the vertical dimension, i.e., using the variations in the local atmospheric pressure—a key component in the positioning of users and HAR. However, since this is a relatively complex usage of barometers compared to the measurement of Earth surface elevation, this local atmospheric pressure could be affected by numerous factors (see Table 3).

From a fundamental point of view, these factors include the atmosphere's fluid properties, such as density, moisture content, temperature, and motion. Speaking more generally, we can attribute the change in local barometric pressure to more abstract factors, such as climate and weather [116–118], air velocity during motion, altitude, and built environment, with intricate interplays between all these factors. The magnitudes of the resultant effects vary and the corresponding ranges can be quantified for specific circumstances. Several studies have attempted to quantify these effects and provide a rich understanding of the factors that influence barometric pressure. Based on those studies, the magnitude of changes and the corresponding timescales for the factors influencing barometric pressure are summarized in Figure 1.

Table 3. Literature related to factors affecting barometric pressure.

Factors	References
Climate and weather	[28,68,69,78,79,85,99,100,104,108,116–118]
Built environment	[66,68,74,84,101,119–121]
Air velocity during motion	[77–79,106,122–124]
Sensor accuracy	[66,69,72,76,78,85,90,99,100,108,113]

In order to understand and demonstrate the effects of the numerous factors that may affect the barometric pressures, we collected a series of experimental data using a custom-made device embedded with a barometer sensor (BMP280). The device was designed specifically for national-scale human activity data collection and analysis [125]. In addition to this device, we also used two mobile phones (Iphone 7 and Iphone XR) with the same mobile application (Physics Toolbox Sensor Suite App version 1.3.5) and barometer model (BMP280) in some of the demonstrations for comparison purposes. The results of the collected data are presented in the next subsections.

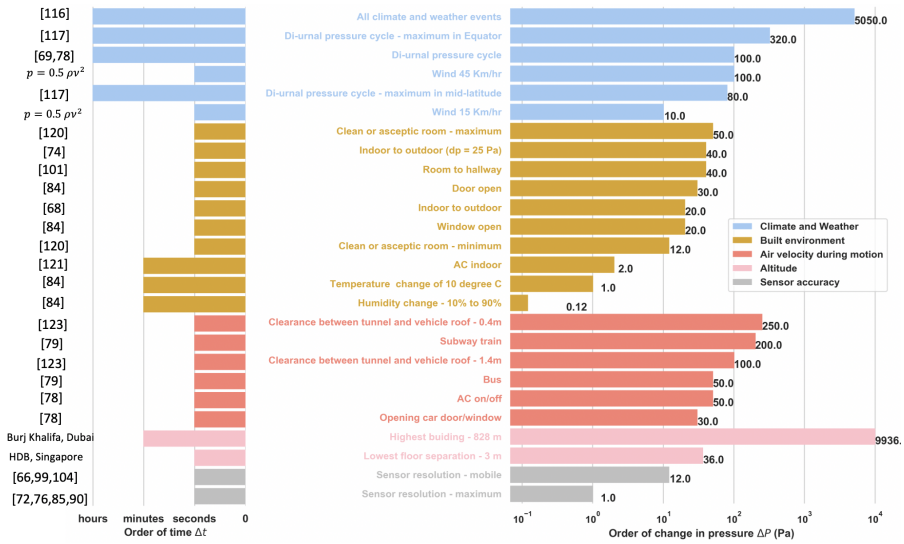


Figure 1. Orders of magnitude of changes in pressure and of the corresponding timescales for several factors influencing barometric pressure.

4.1. Climate and Weather

The atmospheric pressure distribution on the planet is caused by the differential heating of the sun at different latitudes, which varies from the tropics to the polar regions [116]. The Earth's tilt also contributes to the heating difference, as well as the nature of the surface being heated, i.e., ocean or land [116]. Moisture content in the air greatly affects the pressure distribution, as the dry and moist air are heated differently. The differential heating hence produces high- and low-pressure regions on the planet. Due to seasonal differences in heating, the corresponding pressure distribution also changes seasonally [116].

The pressure distribution creates isobaric contours—i.e., the points along which atmospheric pressure is constant [116]. By a combination of the pressure differences in these isobaric regions and the Coriolis force generated by the Earth's rotation, the air is moved from one place to another, creating wind. The resulting forces impact various scales, creating both the long-term climatic system of the planet and the short-term weather patterns observed locally.

It is very important to note that all weather patterns observed are caused by very small changes in pressure—i.e., the maximum change in the sea-level atmospheric pressure never exceeds 5% of the absolute atmospheric pressure [116]. A gentle breeze (15 km/hr) creates a pressure difference of 10 Pa and a strong breeze (45 km/hr) creates a 100 Pa difference [126]. Although the magnitude of the wind factor is significant, their time scales are often slower than the scale of the human activity to be predicted. During the estimation of floor height, Liu et al. [104] showed that a windy day produces error in their estimation of floor height, while Sankaran et al. [79] found that a windy day produced no significant change in their prediction of movement. Sanakaran et al. [79] further noted that neither wind nor rain had any significant impact on HAR. Similarly, Vanini et al. [28], whileperforming their VDA classification, found that neither cloudy nor rainy weather had any impact whatsoever. Bao et al. [85] estimated the error caused by the wind and developed a model that takes into account the dynamic pressure change to remove this effect based on the wind speed. Figure 2a shows an example of the effect of wind on barometer data: The pressure fluctuations when the sensor is exposed to wind are slightly but not significantly more erratic than the noise due to the sensor resolution. The data were collected using one of the custom-built devices [125] equipped with a barometer (BMP280, operating at about 1 Hz sampling rate). The device was placed behind a glass shield in an outdoor environment for

the first hour and then exposed to the incoming wind by placing the device in front of the glass for the next 45 min.

Similarly to surface heating, the upper part of the atmosphere is heated differently over the period of a day. The resulting diurnal temperature variations give rise to a diurnal pressure cycle as illustrated in Figure 2b [117] with data collected with the custom-built device [125] (with BMP280, about 1 Hz sampling rate), which was placed on a desk in an indoor environment for 15 h (from 6:00 a.m. to 8:50 p.m.). This behavior has been well documented since the 1830s [118], with a semi-diurnal cycle with two peaks around 9 a.m.–12 p.m. and 9 p.m.–12 a.m., depending on the latitude [108]. In the tropics, the first late morning/afternoon cycle has the highest amplitude due to the maximum heating during the day, while the second night cycle has a lesser amplitude given the reduced solar heating [117]. In the mid-latitude regions, these peaks are shown to have similar amplitudes for both cycles [117]. The highest pressure variations are reported closer to the equator with 320 Pa between maxima and minima, while the mid-latitude difference does not exceed 80 Pa [117]. In addition to the periodical pressure variation, the absolute pressure also changes from day to day [108]. These changes are, however, prominent in mid-latitude regions (∼600 Pa on average), while they are smaller in the tropics (∼70 Pa on average).

Ho et al. analyzed pressure data from 2309 U.S. cities and observed that these variations are less than 100 Pa per hour for 99% of the time [78], while Liu et al. [69] observed that they could exceed 100 Pa for intervals larger than one hour and remained stable over short intervals of less than one minute. The time scale of the diurnal pressure cycle is hence much larger than many human activity time scales, and these slow variations are not concerning for applications related to study and classification of human behavior [28,100]. Ho et al. [78] used reference pressure from nearby weather stations to alleviate this effect when identifying the correct elevation of a vehicle path. Bao et al. [85] modeled the first-order difference of this pressure cycle as a white Gaussian stochastic process. Other studies have focused on using relative pressure changes called differential pressure instead of the absolute pressure [68,99,100].

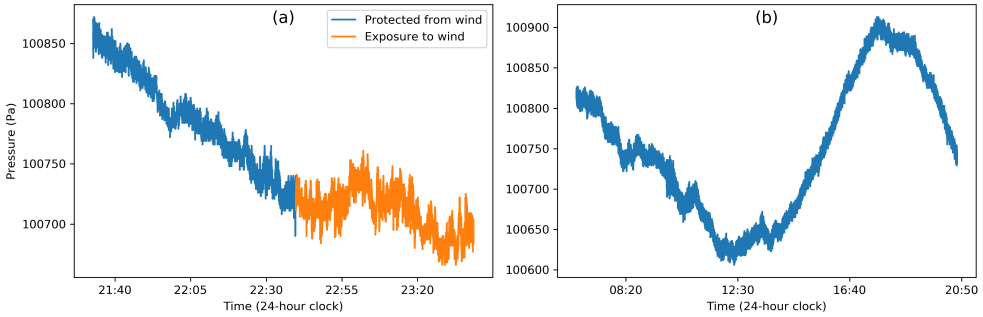


Figure 2. An illustration of the effects of climate and weather on barometric pressure: (**a**) effect of wind and (**b**) diurnal pressure cycle in Singapore. Recorded by a custom-made device (barometer model: BMP280) with ∼1 Hz sampling rate.

4.2. Built Environment

Ambient pressure can be controlled in an enclosed space using mechanical systems like heating, ventilation, and air-conditioning (HVAC) systems [119]. Absolute and relative barometric pressure can thus be affected by these control systems. Buildings are positively pressurized compared to the exterior so as to have an outflow of air [120]. The magnitude of pressure differences is recommended by architects and regulatory authorities based on the function of the space. A minimum of 2.5 Pa is required for general living spaces, while a clean or aseptic isolation room is kept in a pressure difference of 12∼50 Pa to avoid contamination from the outside.

Lstiburek et al. [121] list the four types of pressure field experienced by a building: 1. exterior field—outside a building, 2. interior field—inside a building room, corridors, stairwells, etc., 3. interstitial field—building cavities, and 4. air conveyance system field—air supply, heating, exhaust systems, etc. The temperature, moisture, and pressure inside a building hence fluctuate over the day through the interaction of these pressure fields caused by the building structure, climate and weather, and the mechanical systems inside the structure [121]. Bollmeyer et al. [84] observed that temperature and humidity changes in a room have very little impact on barometric pressure. For instance, a temperature change of 10 degrees Celsius causes a $\pm$1 Pa change, while a humidity change from 10% to 90% creates a pressure difference of less than 0.12 Pa.

Muralidharan et al. [68] observed that the type of building space (tall, short, narrow, and/or wide) and building pressurization also have little effect on barometric pressure measurement. They recorded pressure differences of less than 20 Pa even across multiple days. Xia et al. [66] similarly found no significant effects due to air conditioning in their prediction of floor location.

Even if HVAC systems and the built environment have little effect on HAR, the transition between indoors and outdoors has been shown to produce noticeable pressure changes. Wu et al. [74] found that a constant pressure difference of 25 Pa between the interior and exterior of a building creates a pressure difference of 20$\sim$40 Pa measured by barometers integrated into smartphones when carried during a door opening. Similarly, Bollmeyer et al. [84] found a 30 Pa jump when a door is opened and a 20 Pa jump when a window is opened. Lstiburek et al. [121] showed in 2002 that indoor air conditioning can lead to pressure changes of approximately 2 Pa. More recently, Xu et al. [101] measured pressure differences between a room and the exterior of approximately 40 Pa.

4.3. Air Velocity during Motion

Vehicles like cars and buses are in quasi-equilibrium with their environment due to the vents and ducts allowing the airflow in. This means that the barometric pressure inside a vehicle is very close to the exterior pressure. Note that the term 'exterior' denotes the air pressure in the immediate vicinity of the vehicle and not the ambient pressure far from it, which could be drastically different. This is due to the fact that ambient pressure is increased by the vehicle's motion near its surface where the dynamic pressure increases and the static pressure falls. The stagnation pressure on the surface caused by stopping the airflow near the surface produces a pressure distribution across the vehicle. This, in turn, creates pressure fluctuations inside the vehicle whenever its motion changes. However, this effect is transient, and a quasi-equilibrium is reached quickly between interior and exterior. In order to demonstrate the effect of air velocity on the barometric pressure, we collected the data with five custom-built devices [125] (BMP280, operating at $\sim$0.06 Hz sampling rate) in a bus and a car for $\sim$55 min and $\sim$30 min, respectively. Figure 3a,b illustrate this behavior when a barometer is carried by a human subject during a bus and car ride, respectively, where limited pressure fluctuations ($\sim$50 Pa) are observed.

On the other hand, the transition between a stationary and moving vehicle can be quite noticeable. Ho et al. [78] found that the opening of a door/window during driving creates pressure changes of up to 30 Pa. During a car ride, they showed that switching the air conditioning on and off created a pressure difference of 50 Pa [78]. It is very challenging to attribute it to the climate-control system alone, as the air conditioning also brings in ventilated fresh air from the outside. However, Dimri et al. [77] observed that although there is a range of pressure jumps between different driving conditions (window open/closed, door open/closed, and air conditioning on/off), this did not affect the prediction of whether a vehicle was stationary or in motion. This shows that during vehicle motion, air velocity can substantially determinate the pressure jumps more than the vehicle environment itself. More importantly, the pressure fluctuations experienced by the vehicle during its motion are predominant. This is shown in Figure 3c (collected with a custom-built device [125], operating at $\sim$0.06 Hz sampling rate for $\sim$3 min in a subway train of the Downtown Singapore MRT line,

from Upper Changi station to Jalan Besar station), where successive accelerations and decelerations of the train create significant pressure drops and rises, respectively, when it leaves and approaches a stop.

This effect can be severely amplified by the built environment, such as in tunnels. Vehicles passing through a tunnel experience a "piston effect", where air is pulled inside the tunnel as the air is pushed back by the vehicle in motion [122]. This is shown in Figure 3b, where the car entering a tunnel leads to a drastic change of more than 200 Pa in the pressure measured inside the car. This effect has been shown to create a train of compression waves throughout the tunnel similar to sonic booms [123]. Sankaran et al. [79] showed that the pressure fluctuations during underground subway rides can exceed 200 Pa compared to a bus ride, which shows fluctuations of ∼50 Pa. Barnes et al. [124] analyzed road vehicle passage through tunnels in the Boston metropolitan area and found that the smaller the clearance between the vehicle top and the tunnel roof, the higher the negative pressure drop on the sensor fixed to the tunnel roof, measuring a drop of 100 Pa and 250 Pa for a clearance of 1.4 and 0.4 m respectively. During subway rides, we can clearly distinguish between a train stop and motion [79] (see also Figure 3c). By combining this behavior with relative elevation data of train stations, Hyuga et al. [106] used the pressure jumps to estimate the location of a user during a subway ride.

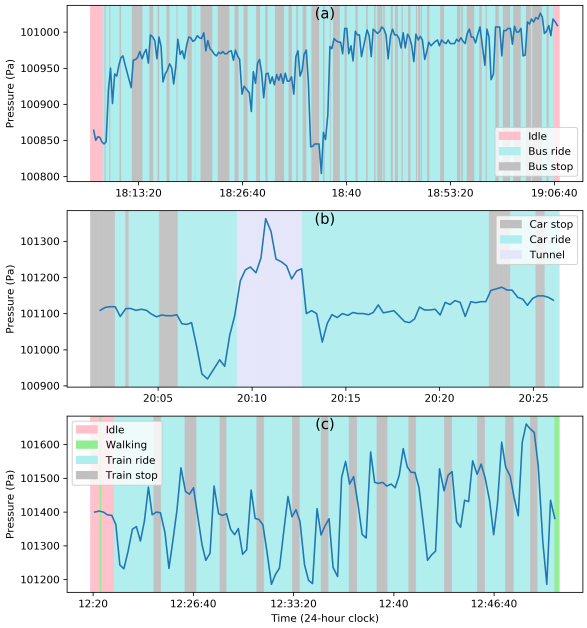


Figure 3. Illustration of the effect of air velocity during motion for a (**a**) bus ride, (**b**) car ride and the effect of the built environment (tunnel), and (**c**) subway train ride. Each panel shows the effect of air velocity during motion on the barometric pressure, with different modes of transport showing different types of changes. Subway train rides yield the highest changes in magnitude (∼200 Pa), while bus and car rides show relatively smaller amplitude changes (∼50 Pa), except when a car is entering a tunnel. The panels also show the effect of elevation changes during travel. In panel (**a**), there is a fall and rise in pressure around 18:35, corresponding to a climb and descent on the road path. Similarly, a drop and rise in pressure is observed in panel (**b**), where the car climbed and descended on an elevated bridge just before entering a tunnel. In panel (**c**), the acceleration of the train as it leaves the stop creates a sudden pressure drop, and a rise in pressure is subsequently observed as it approaches a stop, followed by a small dip in pressure as the train comes to equilibrium with the station environment, thus creating a repeating pressure pattern. Data were recorded by carrying a custom-made device (barometer model: BMP280) in different transport modes with a sampling rate of ∼0.06 Hz.

4.4. Altitude

Atmospheric pressure falls as we travel vertically upward above the Earth's surface. This is due to the Earth's gravitational pull of air molecules to the surface, which gives rise to a pressure gradient equal to $-\rho g$, where ρ is the air density and g is the local acceleration of gravity. As a consequence, the rate of change of altitude with pressure is almost linear near the Earth's surface, while it is almost exponential at higher altitudes when taking into account the variations of the air density with pressure and temperature [115]. This is again due to gravity, which is stronger near the Earth's surface, combined with the fact that air molecules in lower atmosphere are compressed by the air molecules above them [127]. For all purposes involving human activity under 10 km altitude above sea-level, it is safe to assume a linear relationship with altitude that typically shows a pressure decrease of 115 Pa per 10 m climb [115].

To illustrate the magnitude of the effect in the context of human activities, we collected some data using two custom-built devices [125] and two mobile phones (Iphone 7 and Iphone XR) operating at 1 Hz in a building, using different vertical movement modes, including elevator rides, escalator rides, and stair-climbing. Figure 4 shows the barometric pressure change due to change in elevation through the three vertical movement modes. For the elevator ride (Figure 4a), we collected three types of movements between seven floors in a campus building (Singapore University of Technology and Design). First, the elevator was stopped at each individual floor, and the process was repeated twice (two descend and two climb). Then, the elevator was stopped once at the fourth floor between the ground and top floors; this was repeated six times. In the third part, the elevator moved directly from the top to the ground floor without any stop in between for ten times. For the escalator ride (Figure 4b), the data were collected using the same two-way escalators repeatedly and continuously for ten times in a subway train station (Upper Changi MRT station). The floor height difference of the escalator is about 7.3 m. For the stair-climbing (Figure 4c), the data were collected in three interconnected seven-floor buildings (same campus). The interconnections are made possible through bridges running between the third and fifth floors of the three buildings. The first three types of VDA—descent, climbing, and descent—happened within three floors of building 3, followed by walking to building 2 on the first floor, where stairs were climbed from the first to the the seventh (4:40 a.m.) and descended to fifth floor. Using the fifth-floor connection to building 1, the stair case to building 1 was reached and used to climb to the seventh and descend back to the first floor (4:46 a.m.). This was followed by climbing to the second floor of building 2 and descending to the first floor of building 2 using a series of smaller staircases (4:49 a.m.). After this, the following floor jumps were made in building 2: floor 1 to 2, 2 to 5, 5 to 1, 1 to 3, 3 to 1, and 1 to 3, with flat pressure readings indicating walks in corridors. The barometric pressure time series of the three types of VDA indicated that the change in pressure pattern has higher fluctuations as we move from the elevator ride, to the escalator ride, and to stair-climbing. This difference is because of the speed at which the vertical displacement took place and, in the case of stairs, it was affected by the staircase design (type, distance between consecutive stairs, etc.) and walking speed.

Pressure–altitude relationship: The pressure–altitude relationship can be derived from the fundamental equation for fluids at rest [126]. Assuming an incompressible fluid in isothermal conditions, the change in elevation is given by

$$z_2 - z_1 = -\frac{p_2 - p_1}{\gamma}, \tag{1}$$

where $\gamma = \rho g$ is the specific weight of air with density $\rho = 1.225\ \mathrm{kg/m^3}$ and acceleration due to gravity $g = 9.81\ \mathrm{m/s^2}$ at standard sea-level conditions. This pressure–altitude elevation holds with negligible errors as long as the elevation under study is less than 10 km from sea level [126], which is the case in most studies.

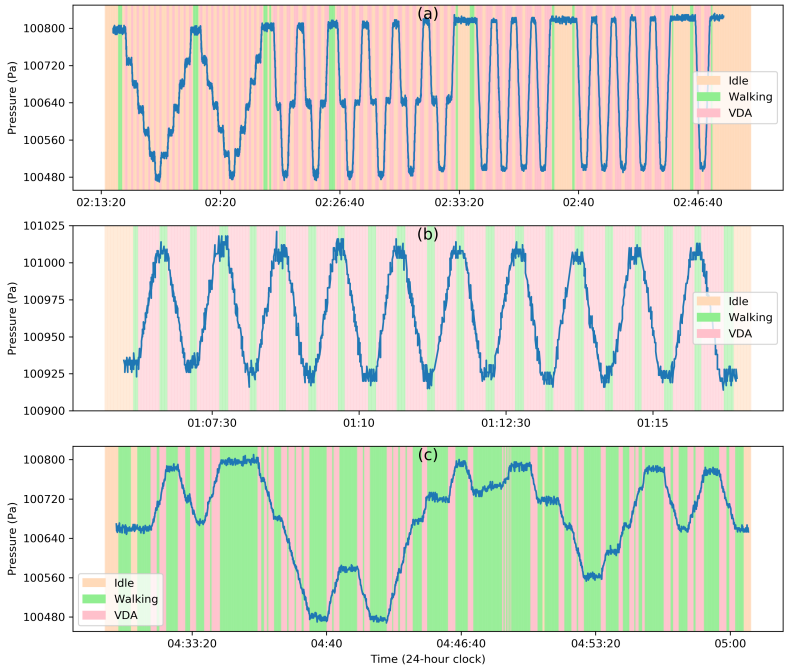


Figure 4. Effects of variations of altitude during human activity and motion on (**a**) an elevator, (**b**) an escalator, and (**c**) stairs. Data were collected by carrying a mobile device with a sampling rate of ∼1 Hz on different modes of vertical movement. For panel (**a**), the elevator was intentionally stopped at each floor in the first part of the data collection (before 2:23 a.m.) to show the ability to distinguish single-floor changes. In the second part (between 2:23 a.m. and 2:33 a.m.), only one stop was made between the first and last floor. Finally, the last part of the data (after 2:33 a.m.) corresponds to an uninterrupted elevator ride between the first and last floor. The data in panel (**b**) were collected by continuously climbing and descending on the same two-way escalator, and the data in panel (**c**) were collected while using stairs. The escalator climb and descent in panel (**b**) correspond to an average relative pressure difference of 80.7 Pa with a standard deviation of 3 Pa, thus showing the accuracy in recording relative pressure changes. Moreover, the relative pressure of 80 Pa corresponds to a height of 6.6 meters according to Equation (1), thus close to the measured height of 7.3 meters and within the equivalence sensor resolution of ±1 meter. The data in panel (**c**) were collected by a human subject performing a series of floor jumps (climbing, descending, and walking) in three interconnected campus buildings: floor 3 to 1 (4:32 a.m.), 1 to 3, 3 to 1, 1 to 7 (4:40 a.m.), 7 to 5, 5 to 7, 7 to 1 (4:46 a.m.), 1 to 2, 2 to 1 (smaller staircases, 4:49 a.m.), 1to 2, 2 to 5, 5 to 1 (4:56 a.m.), 1 to 3, 3 to 1 (4:59 a.m.), and 1 to 3. The barometric pressure fluctuations are highest for the stairs due to the low speed of vertical displacement, difference in staircase designs, and inconsistent walking speed.

4.5. Sensor Accuracy

The quality of measurement of barometric pressure is limited by the sensor's accuracy. Both absolute and relative barometric pressure can change between devices due to differences in sensors and their characteristics. Here, the ensuing measurement errors are defined as being caused by such inherent limitations of the sensor and not due to other factors, such as the environmental ones discussed previously [72,85].

Device dependency: The device dependency is introduced to account for differences between devices and software platforms [66,99,100,108,113], manufacturing inconsistencies, and inappropriate calibration by the manufacturer [100]. Figure 5a illustrates how the time series of barometric pressure

readings from two devices can differ. Using two custom-built devices and two mobile phones (Iphone 7 and Iphone XR) with the same mobile application (Physics Toolbox Sensor Suite App version 1.3.5), a human subject simultaneously carried the devices while climbing and descending on the same two-way escalators for ten times. The recorded barometric pressures between the two custom-built device data shown were significantly different, while the relative pressure was practically constant. Absolute barometric pressure thus needs to be calibrated between different devices for comparison [66]. This can be done before deployment or performed actively by using a reference pressure from nearby weather stations [85,108], buildings, or floor levels [66]. For instance, Ye et al. [99] used active peer-to-peer calibration when users detect each other and used the encounter network to calibrate all the devices. It is thus difficult or impossible to use barometers to measure absolute atmospheric pressure accurately without careful calibration; several studies have shown that it is possible to produce consistent relative pressure measurements [69,99,100,113]. The relative pressure, however, is also affected by the sensor's resolution, drift, and noise.

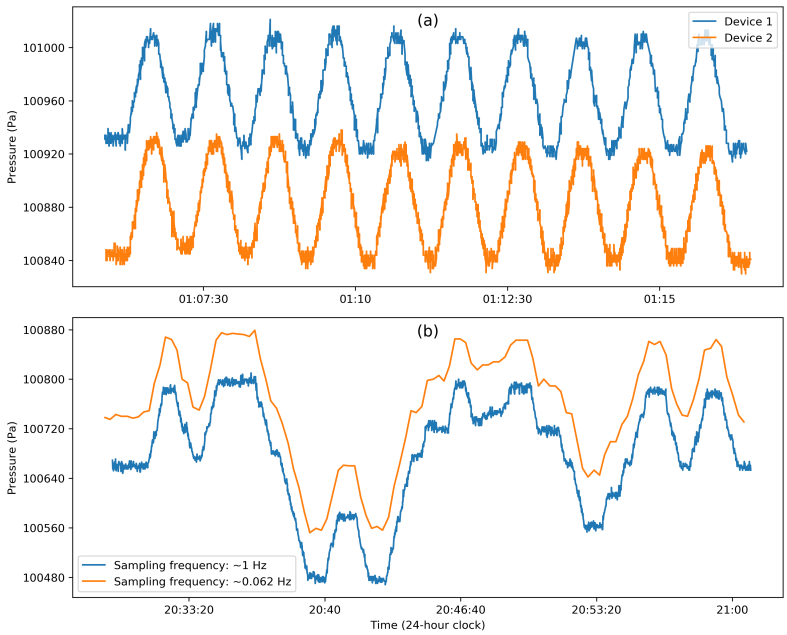


Figure 5. The effects of sensor accuracy. (**a**) Device dependency: Two custom-built devices were simultaneously carried by a human subject on an escalator to record the barometric pressure at a ∼1 Hz sampling rate. The absolute pressure measured by each device is significantly different, while the relative pressure is practically constant. (**b**) Effect of sampling frequency: Two custom-built devices were simultaneously carried by a human subject on stairs. They were both embedded with the same microelectromechanical system (MEMS) barometer model (BMP280), but with different sampling rates, i.e., ∼1 Hz and ∼0.062 Hz, respectively. This panel shows how the sampling rate affects the detection of altitude changes.

Sensor resolution: The accuracy of the barometer is dependent on the built-in resolution of the sensor. Barometers embedded in mobile devices generally have a relative accuracy of ±10 Pa [66,99,104], while commercially available high-resolution sensors can reach an accuracy of ±1 Pa [72,76,85,90]. Sensor resolution is also affected by the measurement errors caused by noise. Haque et al. [72] used Allan Deviation (ADEV), a time-domain analysis, to estimate the non-stationary errors of four different barometer models, and listed the random noise processes that are dominant for a given observation period.

Sensor drift: Some sensors exhibit a drift in time due to faulty manufacturing or old age. Ho et al. [78] found this to be a temporary drift with a non-Gaussian distribution, and were able to remove it by modeling the noise as an Ohnstein–Uhlenbeck diffusion process—a process that pushes the drift towards its mean or center.

Sampling frequency: The recording frequency determines the completeness of the data. Weather stations generally send out data every hour, while mobile barometers embedded in mobile devices can be designed to output at a rate of 1∼20 Hz [76,79,82,83,92,94,104,114] or higher [67,80,89]. Depending on the activity to be recognized, this sampling frequency should be set appropriately to capture the actual time scale of the activity. As an illustration, Figure 5b shows the time series of barometric pressure recorded by two devices with different sampling rates—1 Hz and 0.062 Hz. A human subject carried the two devices using stairs in a seven-floor building. The shift of absolute pressure was caused by the different device issues. Although the overall patterns of the two time series were parallel, some details in the device with the sampling rate of ∼0.062 Hz (orange line) were missing in comparison to the barometric pressure recorded with the ∼1 Hz device. This indicated that altitude or human activity might not be accurately estimated or recognized during certain periods if the sampling rate is not high enough.

5. Discussion

In this review, we had set out to (1) describe sensor data collection processes to track human activities, both in general and more specifically in the case of barometers, (2) discuss the use of the barometer in human activity recognition and list its applications, and (3) understand the many factors that affect barometric pressure. This section reviews the important findings of this endeavor, explores the challenges for each application category, and recommends future research directions.

5.1. Key Findings

In Section 2, we have summarized the road map to sensor data collection by presenting its challenges and nuances. Unlike inertial or location sensors, barometers make use of the subtle changes in the ambient conditions to help track human activities. Depending on the application, barometers are also less dependent on the placement and orientation of the sensor, and of the physical characteristics of the subject themselves. They also require lower sampling rates and less pre-processing than inertial sensors. However, the recognition of activity classes using barometric sensor data faces similar issues to those of other sensors in terms of class imbalance, lack of long-term data in free-living environments, and accurate annotation.

In Section 3, we detailed the common methods used in different stages of activity recognition using barometers and their wide-ranging applications to track human activities. Since the late 1990s, the use of barometers has expanded beyond environmental sensing to tracking human activities. Measuring subtle changes in the ambient pressure has resulted in the ability to detect a vehicle's movement and its location [77–79,106], human gait patterns [87], and vertical mobility [28,65–70], as well as to monitor indoor environments [74]. As listed in Table 1, the limited number of studies that have used only barometers are versatile in their applications, but have mostly considered experiments in specific environments (e.g., focusing only on indoors or on outdoors, on short-term or on long-term periods), limited classes of activities, and, more often than not, have failed to consider all the factors that can potentially affect the barometric pressure. The tracking of activities using barometric pressure data is complemented by geographical elevation maps, usage of multiple barometers, or other sources of barometric pressure data, such as reference stations. However, an accurate inference of a wide range of activities can be done by incorporating additional sensors' data, such as inertial, location, or environmental sensors. In fact, these sensors can be important in removing or quantifying undesired factors that affect barometric pressure; for example, the use of GPS data can give the context of motion in vehicles, and the accelerometer can provide the state of motion or rest during ambulation. Table 2 showed the list of important studies that have used barometers as part of their sensor suite to

improve the detection of vertical mobility and falls, as well as their estimation of energy expenditure in physical activity.

The studies listed in Tables 1 and 2 highlight both the limitations of their experimental conditions and the lack of understanding of several factors that influence the measured ambient pressure. Section 4 hence endeavors to move in this direction by reviewing the fundamental properties of atmospheric pressure and inspecting their interaction with several environmental conditions that arise when a barometer is carried by a human. It also brings together a range of studies from different fields of science and engineering that have contributed to the improvement of the understanding of the factors influencing atmospheric and barometric pressure and to the quantification of their respective magnitudes (see Figure 1). All weather events that have happened on our planet have been observed to produce less than 5050 Pa change in barometric pressure [116]. While a high-speed wind can cause a change in pressure in the order of 100 Pa [126] in a matter of seconds, many of the changes by climate and weather factors happen over a longer time scale, such as the one due to diurnal pressure cycle (Section 4.1). Climate-controlled buildings similarly cause changes in the measured ambient pressure during the transition between indoors and outdoors, and when opening and closing doors or windows (Section 4.2). Even though this causes a transient change in barometric pressure measurement, these changes are observed to be less than 40 Pa [74,84,101,121]. During vehicle motion, the dynamic pressure in the immediate vicinity of the vehicle increases and decreases significantly during acceleration and deceleration, respectively (Section 4.3). This produces a transient change in pressure of 50∼300 Pa in a matter of seconds [79,124]. This change is magnified by the presence of built environment like tunnels and bridges. In addition, the sensor itself can be a source of error. For example, the difference in measurement of absolute pressure between devices, the inherent resolution of the sensors, sensor drift over a long time, and the sampling frequency of the collected data can all add to the noise in the measurement (Section 4.5). The approximately linear relationship between barometric pressure and altitude for less than 10 km from sea level is hence affected by all these factors [126]. Any reasonably long-term tracking of individuals using a barometer should hence be aware of these factors to interpret the collected data accurately.

5.2. Challenges and Future Directions for Different Applications

The various factors that influence barometric pressure have been used by many studies to make better modeling decisions to accurately identify the activity class of interest. The steps to understand these factors have recently led to creative use of the barometric pressure measurements to identify some classes of activities that are traditionally detected using other sensors. To review the challenges and recommend future directions for different applications of barometer data, we categorized the studies that track human activities into the following groups—(1) human mobility: tracking movement of people, (2) health monitoring: tracking elderly, patients, or healthy people for health-related activities, (3) vehicle tracking: tracking movement of people riding vehicles such as cars, buses, trains, etc., and (4) building monitoring: monitoring the movement of people through the changes in building environments. A taxonomy of barometer applications to track human activities is shown in Table 4.

Table 4. Taxonomy of barometer applications to track human activities.

Application	Description	Reference
Human mobility	Tracking movement of people	[28,67–69,84,87,101,104]
Health monitoring	Tracking elderly, patients, or healthy people for health-related activities.	[44,44,75,76,83,91,93,95,97,112]
Vehicle tracking	Tracking movement of people riding vehicles such as cars, buses, trains, etc.	[77–79,106]
Building monitoring	Monitoring the movement of people through the changes in building environments.	[66,68,71,74,84,99,108,121]

Human mobility: The subtle changes in barometric pressure during body movements are used to detect walking, understand gait patterns, and count steps [87]. Similarly to the case of accelerometers, which are typically used in this application, this requires a high sampling rate ($\sim$5 Hz) for barometers (albeit less than the one needed for accelerometers, $\sim$20 Hz). This tracking also requires a high sensor resolution ($\pm$1 Pa) that is usually not present in commonly available devices, like mobile phones ($\pm$12 Pa) [28]. Hence, future tracking of walking and gait patterns requires custom-built devices that have a high resolution and sampling rate, or the wearable industry should move in the direction of assembling such high-performance barometers in ubiquitously available devices. However, this will also lead to higher battery consumption.

In the field of indoor navigation and positioning [28,65–73], vertical displacement activity can be tracked with barometer data. One important challenge in tracking VDA accurately is the need for a reference pressure to keep track of the floor changes. Another is the lack of accounting for factors—other than altitude—that affect barometric pressure in indoor environments, such as the indoor-to-outdoor transition due to climate-controlled buildings, diurnal pressure cycle that slowly drifts the reference pressure, and difference in absolute pressure measurements that necessitates an active calibration between devices. When aiming at the detection of VDA, barometric signal data need to be considered not only to recognize altitude changes, but also to determine the mode of vertical transport (stair-climbing, slope, escalator or elevator rides). In studying epidemic disease propagation, for example, it makes a significant difference to recognize whether the subject is in an elevator (closed space) or on an escalator (open space). The identification of vertical transportation types could be recognized through some pattern recognition of the experimental barometric pressures. There are a lot of useful extensions of barometer sensor applications in various fields that require further exploration.

Understanding human mobility has far-reaching applications in urban planning and economic development [128]. While traditional study of mobility used GPS or other location-based sensors [129], studying vertical mobility requires the use of barometers. The derived vertical displacement from such studies and subsequent statistics of vertical displacements can help better model and predict mobility in a building or at the city scale. The variation of the natural environmental conditions during mobility over the course of the day leads to a complex interplay of factors that influence the barometric pressure. Hence, the application of studying city-scale human mobility in its vertical dimension demands a deeper understanding of all these factors.

Health monitoring: Currently available wearable devices are increasingly designed to track health-related activities. In healthy populations, physical exercises like running, walking, playing, and the rest time between them are estimated with the help of a suite of sensors, including barometers. Barometers have been known to improve the tracking of physical activity, resulting in a better estimation of energy expenditure [44,91,97,112]. While barometers have been used to identify walking or to count the number of steps [87], their use in tracking other physical activities (e.g., running) has not been explored. To track the elderly or patients with health issues, barometers are primarily used to detect falls from wheelchairs, bed, or stairs [76,83,93]. It is important to note that this requires high-resolution barometers whose errors should be smaller than the height of the potential fall.

Vehicle tracking: As mentioned in Section 3.4, the topographical features in our transportation routes can be used to track vehicles. An important challenge here is that air velocity during motion (Section 4.3) can lead to changes in pressure as high as 300 Pa, which can result in an altitude estimation error of 25 meters (12 Pa $\simeq$ 1 m). On the other hand, the change in pressure due to this factor can itself be used to distinguish the vehicle ride from vehicle stops. The identification of vehicle stops, where there are no pressure jumps, has been used successfully as the point to reliably convert the barometric pressure into altitude [106]. A better understanding of the relationship between the barometric pressure and the type of vehicle, velocity of travel, and configurations of the vehicle (e.g., opened/closed windows) could lead to better models and predictions. Moreover, similarly to

tracking VDA over long time, a reference pressure from local stations and active calibration between devices are necessary to accurately keep track of the elevation changes during vehicle rides.

One important application of tracking vehicles is to locate and track the users in lieu of using GPS [44,77] (or other location sensors), which are known to be energy expensive. While this requires an initial location input, reference pressure, active calibration, geographic elevation map, and tracking of vehicle stops (or slowing down), the primary limitations is that it might work only in routes with measurable change in elevation. It is also worth noting that in the absence of an efficient algorithm to match the barometric pressure changes to possible routes using the elevation map, using a barometer might be as consume as much energy as using a GPS. Incidentally, in mountainous regions where mobile network connectivity is usually sparse, this application may particularly find its niche to supplement GPS.

Building monitoring: As mentioned in Section 4.2, buildings with HVAC systems have a dynamic airflow environment where barometric pressure change is noticeable during indoor-to-outdoor transitions and opening/closing of doors or windows. While there are applications for building security [74], this also enables tracking of human movements. For example, the identification of indoor-to-outdoor transitions has been used to begin the calibration of devices entering a building to keep track of floor changes. Inside the buildings, the knowledge of different room pressurization or conditions of open/semi-open spaces can be used to determine the location of the user and help improve the indoor localization and navigation systems in a building. This naturally requires barometers to be carried by the users and installed in the environment. Similarly to other applications, detecting changes in pressure that are typically <40 Pa in indoor environments requires barometers with a high sensor resolution ($\pm$1 pa) and sampling rate ($\sim$5 Hz), which are not available in mobile devices. Hence, an array of barometric sensors to monitor a building can lead to an accurate tracking of human movements in indoor environments.

Furthermore, barometers are increasingly used as a practical complement to other sensors for tracking human activities. As noted before, a multi-sensor setup could provide more flexibility in the recognition of activities and more comprehensive information on the environment of the sensors and carriers. However, a multi-sensory data collection setup would generate a large variety in collected data, leading to challenges in data integration. Methods that enable fusion of sensor data of barometric signals to other sensor data need to evolve to fully realize the barometers' full potential.

Tables 1 and 2 also show that for many applications, not all the factors that can influence pressure are always considered, even after taking into consideration the limited scope of these studies. These factors are not always fully understood, and their magnitude is not universally agreed upon due to the wide range of conditions under which these experiments are performed (see Figure 1). It is thus crucial to understand and quantify all the factors affecting the barometric pressure when working on a particular problem or application related to human activity tracking and recognition. Some of these effects may be irrelevant, while others might impair our ability to properly identify patterns of activity from the sensed data.

5.3. Privacy Issues

Tracking human activities inevitably leads to privacy issues. Anonymity of individuals cannot be guaranteed even with coarse spatial and temporal resolutions of the collected data [130]. Re-identification of individuals from seemingly anonymous data has been shown to be effectively performed due to unique signatures of human movement [130,131]. This issue arises as well with the use of barometers, where re-identification of human subjects could be possible using the tracking of altitude changes during the start and end of each day. In effect, a barometer can be used to perform the so-called last-mile tracking that increases the spatial resolution from a block to a floor. This is a particularly pronounced problem in highly vertical cities, and a proper framework to protect the privacy should take into account the potential use of barometer data when tracking human activity. Indeed, privacy issues are always a major challenge for all human activity analysis or social-related

studies. Previous research introduced a series of geo-masking techniques for the improvement of privacy protection [132–135], but most of these techniques could only be used for 2D spatial data, leaving a large gap for the human activities in the vertical dimension.

6. Conclusions

Today, barometers are found in almost (if not all) wearable devices and smartphones. The vast breadth of applications listed in this review underscores the tremendous potential for use in extracting barometric pressure data on a massive scale. Some attempts with limited scope are reported in Tables 1 and 2. However, the systematic use of barometer data could be envisioned on a large scale and on long time scales as a tool to study some aspects of the behavior of large populations of humans (or other living creatures) in their free-living environments, similarly to what has been done with mobile phone call detail records for studies of communications and movement.

Here, we reviewed and discussed a series of previous studies on using barometers in the tracking and recognition of human activity and movement patterns. Although numerous factors may affect the barometric pressure, the potential of using barometers as sensors for tracking human activity is enormous and worth the anticipation. While most previous studies focused on short-term, indoor, and small-scale human activities, currently, there is a large research gap in the utilization of barometers for a study on a large spatial scale (e.g., city-wide or on a national level) and on a long temporal scale for human activity tracking, recognition, and analysis, specifically concerning the detection of vertical movements.

Author Contributions: Conceptualization: R.B.; methodology: R.B. and A.M.; writing—original draft preparation: R.B. and A.M.; writing—review and editing: R.B., A.B., and W.C.B.C.; supervision: R.B. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a Singapore University of Technology and Design (SUTD) grant (Cities Sector: PIE-SGP-CTRS-1803).

Acknowledgments: We would like to thank Elias J. Willemse for the fruitful discussions and comments. We also would like to thank Bige Tunçer and her research team for their assistance with barometric sensors and guidance on data collection. A.M. is supported by an MOE-SUTD PhD fellowship.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MEMS	Microelectromechanical System
GPS	Global Positioning System
ADL	Activities of Daily Living
UAV	Unmanned Aerial Vehicle
IoT	Internet of Things
IMU	Inertial Measurement Unit
Pa	Pascal (unit)
Hz	Hertz (unit)
PCA	Principal Component Analysis
VDA	Vertical Displacement Activities
SMOTE	Synthetic Minority Over-sampling Technique
BMI	Body Mass Index
ML	Machine Learning
RFID	Radio Frequency IDentification
FIR	Finite Impulse Response filters
IIR	Infinite Impulse Response filters
DP	Differential Pressure

HAR	Human Activity Recognition
SVM	Support Vector Machine
LSTM	Long Short Term Memory networks
HMM	Hidden Markov Model
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
DTW	Dynamic Time Warping
HVAC	Heating, Ventilation, and Air-Conditioning system
AC	Air-Conditioning system
ADEV	Allen Deviation

References

1. Magie, W.F. *A Source Book in Physics*; Harvard University Press: Cambridge, MA, USA, 1969.
2. Pascal, B.; Barry, F.; Spiers, I.H.B.; Spiers, A.G.H.; Stevin, S.; Galilei, G.; Torricelli, E. *The Physical Treatises of Pascal: The Equilibrium of Liquids and the Weight of the Mass of the Air*; Michigan Publishing, University of Michigan Library: Ann Arbor, MI, USA, 1937.
3. Middleton, W.E. *The History of the Barometer*; Baros Books: Trowbridge, UK. 1994.
4. Figuier, L.; Gautier, É. *L'Année Scientifique et Industrielle*; Hachette et Cie: Paris, France, 1861; Volume 5.
5. Five Billion Bosch MEMS Sensors. Available online: https://www.bosch-presse.de/pressportal/de/en/five-billion-bosch-mems-sensors-42857.html (accessed on 15 August 2020).
6. Barometric Pressure Sensors: The Design Engineer's Guide: Avnet Abacus. Available online: https://www.avnet.com/wps/portal/abacus/solutions/technologies/sensors/pressure-sensors/media-types/barometric/ (accessed on 15 August 2020).
7. Lara, O.D.; Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [CrossRef]
8. Del Rosario, M.B.; Redmond, S.J.; Lovell, N.H. Tracking the evolution of smartphone sensing for monitoring human movement. *Sensors* **2015**, *15*, 18901–18933. [CrossRef]
9. Parviainen, J.; Kantola, J.; Collin, J. Differential barometry in personal navigation. In Proceedings of the 2008 IEEE/ION Position, Location and Navigation Symposium, Monterey, CA, USA, 5–8 May 2008; pp. 148–152.
10. Janidarmian, M.; Fekr, A.R.; Radecka, K.; Zilic, Z. A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors* **2017**, *17*, 529, [CrossRef] [PubMed]
11. Twomey, N.; Diethe, T.; Fafoutis, X.; Elsts, A.; McConville, R.; Flach, P.; Craddock, I. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics* **2018**, *5*, 27, [CrossRef]
12. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* **2014**, *14*, 10146–10176. [CrossRef]
13. Chung, S.; Lim, J.; Noh, K.J.; Kim, G.; Jeong, H. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors* **2019**, *19*, 1716, [CrossRef]
14. Willetts, M.; Hollowell, S.; Aslett, L.; Holmes, C.; Doherty, A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Sci. Rep.* **2018**, *8*, 1–10, [CrossRef]
15. Yang, X.; Stewart, K.; Tang, L.; Xie, Z.; Li, Q. A review of GPS trajectories classification based on transportation mode. *Sensors* **2018**, *18*, 3741, [CrossRef]
16. Huang, H.; Cheng, Y.; Weibel, R. Transport mode detection based on mobile phone network data: A systematic review. *Transp. Res. C Emerg. Technol.* **2019**, *101*, 297–312, [CrossRef]
17. Mun, M.; Estrin, D.; Burke, J. Parsimonious mobility classification using GSM and WiFi traces. In Proceedings of the Fifth Workshop on Embedded Networked Sensors (HotEmNets), Raleigh, NC, USA, 5–7 November 2008; pp. 1–5.
18. Sapiezynski, P.; Stopczynski, A.; Gatej, R.; Lehmann, S. Tracking human mobility using WiFi signals. *PLoS ONE* **2015**, *10*, e0130824, [CrossRef]
19. Keramat Jahromi, K.; Zignani, M.; Gaito, S.; Rossi, G.P. Simulating human mobility patterns in urban areas. *Simul. Model. Pract. Theory* **2016**, *62*, 137–156, [CrossRef]

20. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Trans. Big Data* **2016**, *3*, 208–219, [CrossRef]
21. Prelipcean, A.C.; Gidófalvi, G.; Susilo, Y.O. Transportation mode detection–an in-depth review of applicability and reliability. *Transp. Rev.* **2017**, *37*, 442–464, [CrossRef]
22. Kunze, K.; Lukowicz, P. Sensor placement variations in wearable activity recognition. *IEEE Pervasive Comput.* **2014**, *13*, 32–41, [CrossRef]
23. Ustev, Y.E.; Durmaz Incel, O.; Ersoy, C. User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, Zurich, Switzerland, 8–12 September 2013; pp. 1427–1436.
24. Khan, A.M.; Lee, Y.K.; Lee, S.Y.; Kim, T.S. Human activity recognition via an accelerometer-enabled-smartphone using Kernel Discriminant Analysis. In Proceedings of the 2010 5th International Conference on Future Information Technology (FutureTech 2010), Busan, Korea, 21–23 May 2010. pp. 1–6. [CrossRef]
25. Chen, Z.; Zhu, Q.; Soh, Y.C.; Zhang, L. Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3070–3080, [CrossRef]
26. Pannurat, N.; Thiemjarus, S.; Nantajeewarawat, E.; Anantavrasilp, I. Analysis of optimal sensor positions for activity classification and application on a different data collection scenario. *Sensors* **2017**, *17*, 774, [CrossRef]
27. Maurer, U.; Smailagic, A.; Siewiorek, D.P.; Deisher, M. Activity recognition and monitoring using multiple sensors on different body positions. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06), Cambridge, MA, USA, 3–5 April 2006; p. 4.
28. Vanini, S.; Faraci, F.; Ferrari, A.; Giordano, S. Using barometric pressure data to recognize vertical displacement activities on smartphones. *Comput. Commun.* **2016**, *87*, 37–48. [CrossRef]
29. Jerri, A.J. The Shannon sampling theorem—Its various extensions and applications: A tutorial review. *Proc. IEEE* **1977**, *65*, 1565–1596. [CrossRef]
30. Khan, A.; Hammerla, N.; Mellor, S.; Plötz, T. Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recognit. Lett.* **2016**, *73*, 33–40. [CrossRef]
31. Yan, Z.; Subbaraju, V.; Chakraborty, D.; Misra, A.; Aberer, K. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. In Proceedings of the International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 17–24, [CrossRef]
32. Feng, T.; Timmermans, H.J. Transportation mode recognition using GPS and accelerometer data. *Transp. Res. C Emerg. Technol.* **2013**, *37*, 118–130, [CrossRef]
33. Shen, L.; Stopher, P.R. Should we change the rules for trip identification for GPS travel records? In Proceedings of the 36th Australasian Transport Research Forum ATRF, Brisbane, Australia, 2–4 October 2013; pp. 1–11.
34. Münzner, S.; Schmidt, P.; Reiss, A.; Hanselmann, M.; Stiefelhagen, R.; Dürichen, R. CNN-based sensor fusion techniques for multimodal human activity recognition. In Proceedings of the 2017 ACM International Symposium on Wearable Computers, Hawaii, HI, USA, 13–15 September 2017; pp. 158–165.
35. Blanke, U.; Schiele, B. Remember and transfer what you have learned-recognizing composite activities based on activity spotting. In Proceedings of the International Symposium on Wearable Computers (ISWC), Seoul, Korea, 10–13 October 2010; pp. 1–8.
36. Helaoui, R.; Niepert, M.; Stuckenschmidt, H. Recognizing interleaved and concurrent activities: A statistical-relational approach. In Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications (PerCom), Seattle, WA, USA, 21–25 March 2011; pp. 1–9.
37. Cleland, I.; Donnelly, M.P.; Nugent, C.D.; Hallberg, J.; Espinilla, M.; Garcia-Constantino, M. Collection of a Diverse, Realistic and Annotated Dataset for Wearable Activity Recognition. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops 2018), Athens, Greece, 19–23 March 2018; pp. 555–560. [CrossRef]
38. Chen, Y.; Shen, C. Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition. *IEEE Access* **2017**, *5*, 3095–3110, [CrossRef]
39. Plötz, T.; Hammerla, N.Y.; Olivier, P.L. Feature learning for activity recognition in ubiquitous computing. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
40. Garcia-Ceja, E.; Galván-Tejada, C.E.; Brena, R. Multi-view stacking for activity recognition with sound and accelerometer data. *Inf. Fusion* **2018**, *40*, 45–56, [CrossRef]

41. Jordao, A.; Nazare Jr, A.C.; Sena, J.; Schwartz, W.R. Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art. *arXiv preprint* **2018**, arXiv:1806.05226.
42. Murad, A.; Pyun, J.Y. Deep recurrent neural networks for human activity recognition. *Sensors* **2017**, *17*, 2556, [CrossRef]
43. De-La-Hoz-Franco, E.; Ariza-Colpas, P.; Quero, J.M.; Espinilla, M. Sensor-based datasets for human activity recognition–A systematic review of literature. *IEEE Access* **2018**, *6*, 59192–59210. [CrossRef]
44. Wang, Z.; Jiang, M.; Hu, Y.; Li, H. An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 691–699, [CrossRef]
45. Banos, O.; Damas, M.; Pomares, H.; Prieto, A.; Rojas, I. Daily living activity recognition based on statistical feature quality group selection. *Expert Syst. Appl.* **2012**, *39*, 8013–8021, [CrossRef]
46. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. *Esann* **2013**, *3*, 3.
47. Nazabal, A.; Garcia-Moreno, P.; Artes-Rodriguez, A.; Ghahramani, Z. Human Activity Recognition by Combining a Small Number of Classifiers. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 1342–1351. [CrossRef]
48. Lara, Ó.D.; Prez, A.J.; Labrador, M.A.; Posada, J.D. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive Mob. Comput.* **2012**, *8*, 717–729, [CrossRef]
49. Bao, L.; Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. In *International Conference on Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3001, pp. 1–17.
50. Bota, P.; Silva, J.; Folgado, D.; Gamboa, H. A Semi-Automatic Annotation Approach for Human Activity Recognition. *Sensors* **2019**, *19*, 501, [CrossRef] [PubMed]
51. Inoue, M.; Inoue, S.; Nishida, T. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif. Life Robot.* **2018**, *23*, 173–185, [CrossRef]
52. Guan, Y.; Plötz, T. Ensembles of deep LSTM learners for activity recognition using wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–28. [CrossRef]
53. Ermes, M.; Pärkkä, J.; Mäntyjärvi, J.; Korhonen, I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Trans. Inf. Technol. Biomed.* **2008**, [CrossRef] [PubMed]
54. Nguyen, K.T.; Portet, F.; Garbay, C. Dealing with Imbalanced data sets for Human Activity Recognition using Mobile Phone sensors. In Proceedings of the 3rd International Workshop on Smart Sensing Systems, Rome, Italy, 25–28 June 2018.
55. Capela, N.A.; Lemaire, E.D.; Baddour, N.; Rudolf, M.; Goljar, N.; Burger, H. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. *J. Neuroeng. Rehabil.* **2016**, *13*, 1–10, [CrossRef]
56. Almaslukh, B.; Artoli, A.M.; Al-Muhtadi, J. A robust deep learning approach for position-independent smartphone-based human activity recognition. *Sensors* **2018**, *18*, 3726, [CrossRef]
57. Nweke, H.F.; Teh, Y.W.; Alo, U.R.; Mujtaba, G. Analysis of multi-sensor fusion for mobile and wearable sensor based human activity recognition. In Proceedings of the International Conference on Data Processing and Applications, Guangdong, China, 12–14 May 2018; pp. 22–26.
58. Mehrang, S.; Pietilä, J.; Korhonen, I. An activity recognition framework deploying the random forest classifier and a single optical heart rate monitoring and triaxial accelerometer wrist-band. *Sensors* **2018**, *18*, 613, [CrossRef]
59. Stikic, M.; Larlus, D.; Ebert, S.; Schiele, B. Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2521–2537, [CrossRef]
60. Incel, O.D.; Ozgovde, A. ARService: A Smartphone based Crowd-Sourced Data Collection and Activity Recognition Framework. *Procedia Comput. Sci.* **2018**, *130*, 1019–1024, [CrossRef]
61. Anjum, A.; Ilyas, M.U. Activity recognition using smartphone sensors. In Proceedings of the 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2013; pp. 914–919, [CrossRef]
62. Sztyler, T.; Stuckenschmidt, H. On-body localization of wearable devices: An investigation of position-aware activity recognition. In Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications, Sydney, Australia, 14–18 March 2016; pp. 1–9, [CrossRef]

63. Khan, A.M.; Lee, Y.K.; Lee, S.Y.; Kim, T.S. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 1166–1172. [CrossRef]
64. Huynh, D.T.G. Human Activity Recognition with Wearable Sensors. Ph.D. Thesis, Technische Universität, Berlin, Germany, 2008.
65. Lang, C.; Kaiser, S. Classifying Elevators and Escalators in 3D Pedestrian Indoor Navigation Using Foot-Mounted Sensors. In Proceedings of the 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nantes, France, 24–27 September 2018; pp. 1–7.
66. Xia, H.; Wang, X.; Qiao, Y.; Jian, J.; Chang, Y. Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning. *Sensors* **2015**, *15*, 7857–7877. [CrossRef] [PubMed]
67. Kronenwett, N.; Qian, S.; Mueller, K.; Trommer, G.F. Elevator and Escalator Classification for Precise Indoor Localization. In Proceedings of the 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nantes, France, 24–27 September 2018; pp. 1–8.
68. Muralidharan, K.; Khan, A.J.; Misra, A.; Balan, R.K.; Agarwal, S. Barometric phone sensors: More hype than hope! In Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, Santa Barbara, CA, USA, 26–27 February 2014; p. 12.
69. Liu, M.; Li, H.; Wang, Y.; Li, F.; Chen, X. Double-Windows-Based Motion Recognition in Multi-Floor Buildings Assisted by a Built-In Barometer. *Sensors* **2018**, *18*, 1061. [CrossRef] [PubMed]
70. Elhoushi, M.; Georgy, J.; Wahdan, A.; Korenberg, M.; Noureldin, A. Using portable device sensors to recognize height changing modes of motion. In Proceedings of the 2014 IEEE International Instrumentation and Measurement Technology Conference (I2mtc), Montevideo, Uruguay, 12–15 May 2014; pp. 477–481.
71. Pipelidis, G.; Rad, O.R.M.; Iwaszczuk, D.; Prehofer, C.; Hugentobler, U. A novel approach for dynamic vertical indoor mapping through crowd-sourced smartphone sensor data. In Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, Japan, 18–21 September, 2017; pp. 1–8.
72. Haque, F.; Dehghanian, V.; Fapojuwo, A.O.; Nielsen, J. A Sensor Fusion-Based Framework for Floor Localization. *IEEE Sensors J.* **2018**, *19*, 623–631. [CrossRef]
73. Shen, X.; Chen, Y.; Zhang, J.; Wang, L.; Dai, G.; He, T. BarFi: Barometer-aided Wi-Fi floor localization using crowdsourcing. In Proceedings of the 2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems, TX, Dallas, USA, 19–22 October 2015; pp. 416–424.
74. Wu, M.; Pathak, P.H.; Mohapatra, P. Monitoring building door events using barometer sensor in smartphones. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 9–11 September 2015; pp. 319–323.
75. Leuenberger, K.; Gonzenbach, R.; Wiedmer, E.; Luft, A.; Gassert, R. Classification of stair ascent and descent in stroke patients. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops, Zurich, Switzerland, 16–19 June 2014; pp. 11–16.
76. Bianchi, F.; Redmond, S.J.; Narayanan, M.R.; Cerutti, S.; Lovell, N.H. Barometric pressure and triaxial accelerometry-based falls event detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2010**, *18*, 619–627. [CrossRef] [PubMed]
77. Dimri, A.; Singh, H.; Aggarwal, N.; Raman, B.; Bansal, D.; Ramakrishnan, K. RoadSphygmo: Using barometer for traffic congestion detection. In Proceedings of the 2016 8th International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 5–10 January 2016; pp. 1–8.
78. Ho, B.J.; Martin, P.; Swaminathan, P.; Srivastava, M. From pressure to path: Barometer-based vehicle tracking. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, Korea, 4–5 November 2015; pp. 65–74.
79. Sankaran, K.; Zhu, M.; Guo, X.F.; Ananda, A.L.; Chan, M.C.; Peh, L.S. Using mobile phone barometer for low-power transportation context detection. In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, Memphis, TN, USA, 3–6 November 2014; pp. 191–205.
80. Sagawa, K.; Ishihara, T.; Ina, A.; Inooka, H. Classification of human moving patterns using air pressure and acceleration. In Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON'98 Cat. No. 98CH36200), Aachen, Germany, 31 August–4 September 1998; Volume 2, pp. 1214–1219.

81. Bolanakis, D.E. Evaluating performance of MEMS barometric sensors in differential altimetry systems. *IEEE Aerosp. Electron. Syst. Mag.* **2017**, *32*, 34–39. [CrossRef]
82. Del Rosario, M.B.; Wang, K.; Wang, J.; Liu, Y.; Brodie, M.; Delbaere, K.; Lovell, N.H.; Lord, S.R.; Redmond, S.J. A comparison of activity classification in younger and older cohorts using a smartphone. *Physiol. Meas.* **2014**, *35*, 2269. [CrossRef]
83. Wang, C.; Lu, W.; Narayanan, M.R.; Chang, D.C.W.; Lord, S.R.; Redmond, S.J.; Lovell, N.H. Low-power fall detector using triaxial accelerometry and barometric pressure sensing. *IEEE Trans. Ind. Inform.* **2016**, *12*, 2302–2311. [CrossRef]
84. Bollmeyer, C.; Esemann, T.; Gehring, H.; Hellbrück, H. Precise indoor altitude estimation based on differential barometric sensing for wireless medical applications. In Proceedings of the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, MA, USA, 6–9 May 2013; pp. 1–6.
85. Bao, X.; Xiong, Z.; Sheng, S.; Dai, Y.; Bao, S.; Liu, J. Barometer measurement error modeling and correction for UAH altitude tracking. In Proceedings of the 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 3166–3171.
86. Sabatini, A.; Genovese, V. A sensor fusion method for tracking vertical velocity and height based on inertial and barometric altimeter measurements. *Sensors* **2014**, *14*, 13324–13347. [CrossRef]
87. Ghimire, B.; Nickel, C.; Seitz, J. Pedestrian motion state classification using pressure sensors. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares, Spain, 4–7 October 2016; pp. 1–6.
88. Voleno, M.; Redmond, S.J.; Cerutti, S.; Lovell, N.H. Energy expenditure estimation using triaxial accelerometry and barometric pressure measurement. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 5185–5188.
89. Moncada-Torres, A.; Leuenberger, K.; Gonzenbach, R.; Luft, A.; Gassert, R. Activity classification based on inertial and barometric pressure sensors at different anatomical locations. *Physiol. Meas.* **2014**, *35*, 1245. [CrossRef]
90. Massé, F.; Gonzenbach, R.R.; Arami, A.; Paraschiv-Ionescu, A.; Luft, A.R.; Aminian, K. Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients. *J. Neuroeng. Rehabil.* **2015**, *12*, 72. [CrossRef]
91. el Achkar, C.M.; Lenoble-Hoskovec, C.; Paraschiv-Ionescu, A.; Major, K.; Büla, C.; Aminian, K. Instrumented shoes for activity classification in the elderly. *Gait Posture* **2016**, *44*, 12–17. [CrossRef] [PubMed]
92. Figueira, C.; Matias, R.; Gamboa, H. Body Location Independent Activity Monitoring. In Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, Rome, Italy, 21–23 February 2016; pp. 190–197.
93. Ejupi, A.; Galang, C.; Aziz, O.; Park, E.J.; Robinovitch, S. Accuracy of a wavelet-based fall detection approach using an accelerometer and a barometric pressure sensor. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 2150–2153.
94. Monteiro, M.; Martí, A.C. Using smartphone pressure sensors to measure vertical velocities of elevators, stairways, and drones. *arXiv preprint* **2016**, arXiv:1607.00363.
95. Nam, Y.; Park, J.W. Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 420–426. [PubMed]
96. Loh, D.; Lee, T.J.; Zihajehzadeh, S.; Hoskinson, R.; Park, E.J. Fitness activity classification by using multiclass support vector machines on head-worn sensors. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 502–505.
97. Anastasopoulou, P.; Tansella, M.; Stumpp, J.; Shammas, L.; Hey, S. Classification of human physical activity and energy expenditure estimation by accelerometry and barometry. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 6451–6454.
98. Kim, S.S.; Kim, J.W.; Han, D.S. Floor detection using a barometer sensor in a smartphone. In Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, Japan, 18–21 September 2017.

99. Ye, H.; Gu, T.; Tao, X.; Lu, J. Scalable floor localization using barometer on smartphone. *Wirel. Commun. Mob. Comput.* **2016**, *16*, 2557–2571. [CrossRef]
100. Zhao, F.; Luo, H.; Zhao, X.; Pang, Z.; Park, H. HYFI: Hybrid floor identification based on wireless fingerprinting and barometric pressure. *IEEE Trans. Ind. Inform.* **2015**, *13*, 330–341. [CrossRef]
101. Xu, Z.; Wei, J.; Zhu, J.; Yang, W. A robust floor localization method using inertial and barometer measurements. In Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, Japan, 18–21 September 2017; pp. 1–8.
102. Kamiya, Y.; Gu, Y.; Kamijo, S. Indoor Positioning in Large Shopping Mall with Context based Map Matching. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 10–12 January 2019; pp. 1–6.
103. Yi, S.; Mirowski, P.; Ho, T.K.; Pavlovic, V. Pose invariant activity classification for multi-floor indoor localization. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 3505–3510.
104. Liu, G.; Iwai, M.; Tobe, Y.; Matekenya, D.; Hossain, K.M.A.; Ito, M.; Sezaki, K. Beyond horizontal location context: Measuring elevation using smartphone's barometer. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, WA, USA, 13–17 September 2014; pp. 459–468.
105. Kaiser, S.; Lang, C. Detecting elevators and escalators in 3d pedestrian indoor navigation. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares, Spain, 4–7 October 2016; pp. 1–6.
106. Hyuga, S.; Ito, M.; Iwai, M.; Sezaki, K. Estimate a user's location using smartphone's barometer on a subway. In Proceedings of the 5th International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments, Seattle, WA, USA, 3 November 2015; p. 2.
107. Wannenburg, J.; Malekian, R. Physical activity recognition from smartphone accelerometer data for user context awareness sensing. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 3142–3149. [CrossRef]
108. Li, B.; Harvey, B.; Gallagher, T. Using barometers to determine the height for indoor positioning. In Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Montbeliard, France, 28–31 October 2013; pp. 1–7.
109. Pipelidis, G.; Moslehi Rad, O.; Iwaszczuk, D.; Prehofer, C.; Hugentobler, U. Dynamic Vertical Mapping with Crowdsourced Smartphone Sensor Data. *Sensors* **2018**, *18*, 480. [CrossRef]
110. Tachikawa, M.; Maekawa, T.; Matsushita, Y. Predicting location semantics combining active and passive sensing with environment-independent classifier. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 220–231.
111. Son, Y.; Oh, S. A barometer-IMU fusion method for vertical velocity and height estimation. In Proceedings of the 2015 IEEE SENSORS, Busan, Korea, 1–4 November 2015; pp. 1–4.
112. Ohtaki, Y.; Susumago, M.; Suzuki, A.; Sagawa, K.; Nagatomi, R.; Inooka, H. Automatic classification of ambulatory movements and evaluation of energy consumptions utilizing accelerometers and a barometer. *Microsyst. Technol.* **2005**, *11*, 1034–1040, [CrossRef]
113. Ichikari, R.; Ruiz, L.C.M.; Kourogi, M.; Kurata, T.; Kitagawa, T.; Yoshii, S. Indoor floor-level detection by collectively decomposing factors of atmospheric pressure. In Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Banff, AB, Canada, 13–16 October 2015; pp. 1–11.
114. Li, Y.; Gao, Z.; He, Z.; Zhang, P.; Chen, R.; El-Sheimy, N. Multi-sensor multi-floor 3D localization with robust floor detection. *IEEE Access* **2018**, *6*, 76689–76699. [CrossRef]
115. Atmospheric Pressure. 2019. Available online: https://www.britannica.com/science/atmospheric-pressure (accessed on 20 November 2019).
116. Cenedese, C.; Smith, P.J. Atmospheric Pressure and Wind. 2019. Available online: https://www.britannica.com/science/climate-meteorology/Atmospheric-pressure-and-wind (accessed on 20 January 2020).
117. Le Blancq, F. Diurnal pressure variation: The atmospheric tide. *Weather* **2011**, *66*, 306–307, [CrossRef]
118. Giles, B. The atmospheric tide: An historical perspective. *Weather* **2012**, *67*, 51–53, [CrossRef]
119. Trane. Commercial Building Pressurization. 2002. Available on: https://www.airtight.ai/wp-content/uploads/2019/07/Commercial-Building-Pressurization-by-Trane.pdf (accessed on 26 November 2020).

120. Hendiger, J.; Chludzińska, M.; Ziętek, P. Influence of the pressure difference and door swing on heavy contaminants migration between rooms. *PLoS ONE* **2016**, *11*, e0155159. [CrossRef] [PubMed]
121. Lstiburek, J.; Pressnail, K.; Timusk, J. Air pressure and building envelopes. *J. Therm. Envel. Build. Sci.* **2002**, *26*, 53–91. [CrossRef]
122. Zhang, N.; Zhou, D. Numerical Analysis of the Pressure Variation in Subway Tunnel When an On-Fire Train Runs at Different Speed. In Proceedings of the 2016 World Congress on Advances in Civil, Environmental, and Materials Research (ACEM16), Jeju Island, Korea, 28 August–1 September 2016; pp. 1–6.
123. Reinke, P.; Flueckiger, M.; Wicht, T. Aerodynamics and Ventilation in Rail Tunnels. May 2015. Available online: https://www.tunneltalk.com/TunnelTECH-May2015-Aerodynamics-and-ventilation-in-rail-tunnels-civil-measures.php (accessed on 26 November 2020).
124. Barnes, J.D.; Brush, E.R.; Newmark, M.S.; Ungar, E.E. Dynamic Pressures on Tunnel Roofs due to Vehicle Passages. *Sound Vib.* **2018**, *52*, 6–8. [CrossRef]
125. Wilhelm, E.; Siby, S.; Zhou, Y.; Ashok, X.J.S.; Jayasuriya, M.; Foong, S.; Kee, J.; Wood, K.L.; Tippenhauer, N.O. Wearable environmental sensors and infrastructure for mobile large-scale urban deployment. *IEEE Sens. J.* **2016**, *16*, 8111–8123. [CrossRef]
126. Young, D.F.; Munson, B.R.; Okiishi, T.H.; Huebsch, W.W. *A Brief Introduction to Fluid Mechanics*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
127. Why Does Atmospheric Pressure Change with Altitude? Available online: https://www.npl.co.uk/resources/q-a/atmospheric-altitude-pressure-changes (accessed on 20 January 2020).
128. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [CrossRef]
129. Barbosa, H.; Barthelemy, M.; Ghoshal, G.; James, C.R.; Lenormand, M.; Louail, T.; Menezes, R.; Ramasco, J.J.; Simini, F.; Tomasini, M. Human mobility: Models and applications. *Phys. Rep.* **2018**, *734*, 1–74. [CrossRef]
130. De Montjoye, Y.A.; Hidalgo, C.A.; Verleysen, M.; Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Sci. Rep.* **2013**, *3*, 1–5, [CrossRef] [PubMed]
131. Zang, H.; Bolot, J. Anonymization of location data does not work: A large-scale measurement study. In Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, Las Vegas, NV, USA, 19–23 September 2011; pp. 145–156.
132. Armstrong, M.P.; Rushton, G.; Zimmerman, D.L. Geographically masking health data to preserve confidentiality. *Stat. Med.* **1999**, *18*, 497–525. [CrossRef]
133. Kwan, M.P.; Casas, I.; Schmitz, B. Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks? *Cartogr. Int. J. Geogr. Inf. Geovis.* **2004**, *39*, 15–28. [CrossRef]
134. Armstrong, M.P.; Ruggles, A.J. Geographic information technologies and personal privacy. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2005**, *40*, 63–73. [CrossRef]
135. Gao, S.; Rao, J.; Liu, X.; Kang, Y.; Huang, Q.; App, J. Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users. *J. Spat. Inf. Sci.* **2019**, *2019*, 105–129. [CrossRef]

5

A Framework for the Identification of Human Vertical Displacement Activity Based on Multi-Sensor Data

This chapter includes a peer-reviewed paper published in IEEE Sensors Journal, which addresses the second research question (R2) posed in Chapter 1, focusing on developing a machine-learning framework to derive vertical displacement based on the insights developed in Chapter 4.

Publication details:

- Manivannan, A., Willemse, E. J., Chin, W. C. B., Zhou, Y., Tunçer, B., Barrat, A., & Bouffanais, R. (2022). A framework for the identification of human vertical displacement activity based on multi-sensor data. IEEE Sensors Journal, 22(8), 8011-8029.

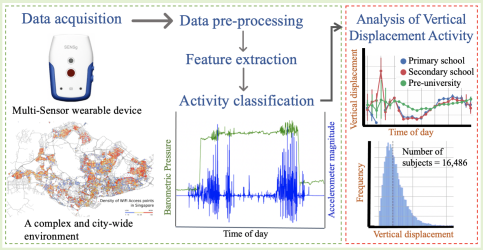
Summary: This paper presents a novel framework for identifying human vertical displacement activity (VDA) using multi-sensor data from a barometer, an accelerometer, and a Wi-Fi scanner. A key contribution of the study is the development of a machine-learning-based framework that effectively distinguishes between vertical and horizontal displacement activities with high accuracy (98% overall accuracy). By applying this framework to a large-scale dataset from over 16,000 participants in Singapore's 2016 National Science Experiment, the paper demonstrates the potential of multi-sensor fusion in improving the precision of vertical mobility tracking. This research advances the understanding of human mobility in densely urbanized environments, particularly in the context of high-rise buildings and vertical cities.

Another significant contribution of the study is its exploration of the factors influencing the accuracy of barometric pressure measurements when used to track vertical movement. The paper addresses key challenges such as atmospheric pressure variations due to weather, built environments, and air velocity during movement, which can all introduce noise into barometric data. We develop several preprocessing techniques to mitigate these effects, thereby improving the reliability of barometer-based vertical displacement detection. This research has far-reaching implications for urban design, indoor navigation, and human activity recognition, especially in the context of smart city applications and vertical urban planning.

A Framework for the Identification of Human Vertical Displacement Activity Based on Multi-Sensor Data

Ajaykumar Manivannan, Elias J. Willemse, Balamurali B. T., Wei Chien Benny Chin, Yuren Zhou, Bige Tunçer, Alain Barrat, and Roland Bouffanais, *Member, IEEE*

***Abstract*—To date, the methodology to track and identify vertical movement from large-scale unstructured data sets is lacking. Here, we design and develop such a framework to accurately and systematically identify the sparse human vertical displacement activity typically buried into the predominantly horizontal mobility. Our framework uses sensor data from a barometer, accelerometer, and Wi-Fi scanner coupled with an extraction step involving a combination of feature engineering and data segmentation. This methodology is subsequently integrated into a machine-learning-based classifier to automatically distinguish vertical displacement activity—with 98% overall accuracy and a 92% F1-score—from its horizontal counterpart. We illustrate the potential of this framework by applying it to an unstructured large-scale data set associated with over 16,000 participants going about their daily activity in the city-state of Singapore. With the vertical movements of this large group uncovered, we can analyze the specific features of this activity class using its statistical distribution. This new knowledge would have significant ramifications for the architectural design of vertical cities.**



***Index Terms*— Wearable sensors, barometer, multi-sensor identification, human activity recognition, vertical displacement activity, stair climbing.**

Manuscript received January 27, 2022; revised February 24, 2022; accepted March 6, 2022. Date of publication March 8, 2022; date of current version April 14, 2022. This work was supported in part by the Singapore National Research Foundation (NRF), in part by the Ministry of Education (MOE), and in part by the Singapore University of Technology and Design (SUTD)-Massachusetts Institute of Technology (MIT) International Design Centre through National Science Experiment (NSE) 2016. The work of Ajaykumar Manivannan was supported by the Ministry of Education (MOE)-Singapore University of Technology and Design (SUTD) Ph.D. Fellowship and currently supported by the University of Ottawa Ph.D. Admission Scholarship. The associate editor coordinating the review of this article and approving it for publication was Prof. Shih-Chia Huang. *(Corresponding author: Roland Bouffanais.)*

Ajaykumar Manivannan and Roland Bouffanais are with the Department of Mechanical Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: roland.bouffanais@uottawa.ca).

Elias J. Willemse is with Waste Labs Pte. Ltd., Singapore 059915.

Balamurali B. T. is with the Engineering Systems and Design (ESD), Singapore University of Technology and Design (SUTD), Singapore 487372.

Wei Chien Benny Chin and Yuren Zhou are with the Engineering Product Development (EPD), Singapore University of Technology and Design (SUTD), Singapore 487372.

Bige Tunçer is with the Architecture and Sustainable Design (ASD), Singapore University of Technology and Design (SUTD), Singapore 487372.

Alain Barrat is with CNRS, CPT, Aix Marseille Univ, Université de Toulon, 13009 Marseille, France, and also with the Tokyo Tech World Research Hub Initiative (WRHI), Tokyo Institute of Technology, Yokohama 226-8503, Japan.

Digital Object Identifier 10.1109/JSEN.2022.3157806

I. Introduction

THE urbanization of our planet is rapidly increasing, with 55% of the world population now living in cities [1]. In 2030, this number is projected to increase to 60% [1]. In the face of this unabated urbanization trend, cities struggle to accommodate the population influx through urban sprawl alone, primarily because of land scarcity and the induced strain on transportation networks. An alternative to urban sprawl currently predominant in the rapidly urbanizing Asia is increasing the density of the built environment, inevitably leading to cities with vertically dominated landscapes and singular skylines.

Such vertical cities exhibit a very distinct urban landscape, manifesting a sprawl of an upward nature with a very high density of high-rise buildings—not necessarily limited to skyscrapers. This vertical growth of cities is reflected in the increasing market demand around the world for vertical transportation systems like elevators and escalators, with approximately 100,000 units installed in 2019 and a forecast for 250,000 units commissioned in 2024 alone [2]. The Asia-Pacific region, where most of the fastest-growing cities in the world are present [1], is said to have the highest growth in demand (85%) for vertical transportation systems [2].

Over the last half-century, today's urban planning has dramatically benefited from extensive human mobility studies. Over the last two decades, this area of research has experienced significant growth due to the convergence of several technological factors: (1) the development of new sensors enabling more accurate tracking of human mobility, (2) the very rapid and massive adoption of mobile phones globally, and (3) the so-called "Big Data" effect. In addition, complexity scientists have developed several new frameworks to analyze and identify specific mobility patterns. Those patterns hidden in troves of high-resolution mobility data have been uncovered thanks to large-scale experiments or from massive commercial databases—e.g., call detailed record (CDR) [3]. However, it is essential to note that these human mobility studies are limited to horizontal movements, i.e., based on a two-dimensional representation of the urban landscape. Roads, railway networks, and pedestrian pathways are modeled on a planar surface in these studies.

As already mentioned, the rapid change in the topology of cities in the developing world, especially in Asia, is prominently three-dimensional [4]. Therefore, the development of sustainable and livable cities heavily depends on studies of human mobility across all three dimensions. However, while brand-new skyscrapers are being erected every day globally, we know surprisingly little about vertical human mobility. For instance, some recent studies propose to study vertical displacements in vertically integrated mixed-use developments as a means to identify key spatial connectors that have a direct influence on social interactions [5]. Moreover, a systematic study of vertical human mobility would benefit urban/infrastructure planning in many ways: e.g., targeted facility allocation in high-rise buildings, optimal placement of vertical nodes based on the typology of the building and the estimated vertical transportation load, effective vertical integration of a building in its neighborhood, etc. [6].

Horizontal mobility has been extensively studied using a wide variety of data sources: e.g., census data, travel surveys, CDR, location-based social network services, GPS [7], and smart travel/transit cards [8]. However, none of these approaches and sensors can effectively track vertical displacements. Interestingly, the sensor technology required to track such human vertical mobility accurately is readily available. What is missing is a methodology that enables the accurate identification of various types of possible vertical displacements from the output of large-scale human experiments with sufficient statistical significance.

Here, we report a contribution towards that goal by introducing and validating a methodology to accurately and systematically identify the sparse human vertical displacement activity (VDA) [9] that is deeply embedded within the predominantly horizontal displacement activity. This methodology is then integrated into a machine-learning-based classifier capable of dealing with large-scale data sets collected in free-living and unstructured urban environments. Classically, barometers have been the primary type of sensor used to track motion in the vertical direction. Indeed, barometric pressure—possibly augmented by other sensors—is commonly used in the field of Human Activity Recognition (HAR) to recognize the particular VDA class, which is of prime interest to us [9]. Specifically, VDA is a particular human activity class that deals with the vertical displacement of individuals in the built environment through commonly available modes of vertical mobility such as stairs, escalators, elevators, or slopes. In this work, the term VDA is intended to encompass human movements in vertically built structure solely. That means we are discarding changes in elevation associated with any vehicle motion (motor vehicle, train, bicycle, cable car, etc.).

Our methodological advancement is thoroughly tested and validated using a big data set obtained from a large-scale human experiment carried out in Singapore: the so-called National Science Experiment (NSE). The NSE was a city-scale experiment that involved 50,000 students in Singapore between 2015 and 2017. The wearable devices designed explicitly for this large-scale experiment were carried by students continuously for five days and contained several sensors, including a barometer and an accelerometer. By fully understanding the complex interplay of factors that influence barometric pressure, we develop several preprocessing methods to alleviate the effects of those factors, with the end goal of achieving the highest possible accuracy in the VDA identification process. Moreover, the VDA extraction process must be robust enough to handle inherent limitations associated with such large-scale human experiments—i.e., low sampling rate, heterogeneity in devices and participant population, missing data, and sensor errors. As part of this process, we manually label a large number of training data (81 subjects, 81 devices, for a time period of 24 hours). This step is followed by a validation using a short-term video-annotated data set (2 subjects, 5 devices, for a period of 6 hours). Finally, we integrated the developed VDA identification methodology into a machine-learning-based classifier. Subsequently, we applied it to the large-scale NSE data set to extract unique features of human vertical mobility associated with the student population participating in the NSE.

The main contributions of this paper can be summarized as follows (see Fig. 1):

- A novel and accurate multi-sensory identification of vertical displacement activity is developed and validated against a sparse data set from a large-scale human experiment involving over 16, 000 individuals going about their daily activity within a densely urban environment.
- The accuracy of this VDA identification process is found to be strongly dependent on a several constraints associated with the sensing of key physical quantities. Specifically, we design a feature extraction step involving a combination of feature engineering and data segmentation. In addition, the properties of the sensors and how they are used in such large-scale experiments create many challenges, which are identified and addressed.
- Using our novel VDA identification process, a machine-learning-based classifier allows us to carry out the first large-scale analysis of human vertical mobility in a city-scale experiment. Interestingly, our results reveal a highly heterogeneous distribution of vertical activity, both in terms of the number of events and of the size of vertical

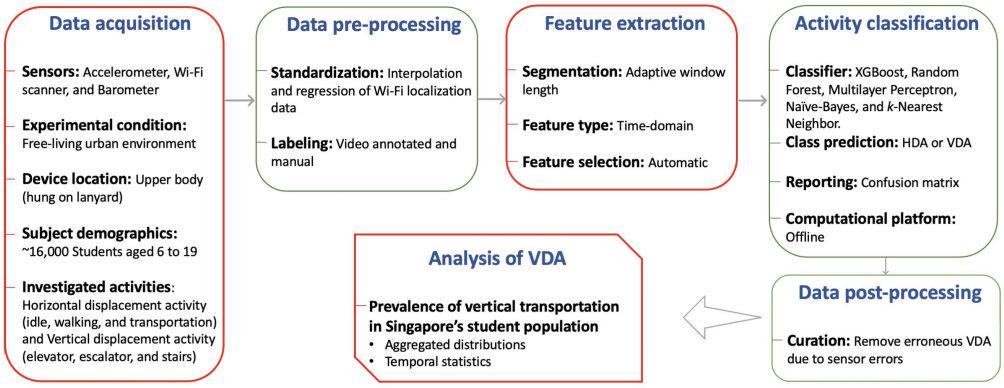


Fig. 1. Human activity recognition (HAR) process in the identification of VDA [10]. This flowchart describes the overall process developed to achieve our VDA analysis represented in the last rectangular box. The key contributions of this work are in the boxes with a red outline.

jumps. These results have far-reaching implications for the architectural design of dense urban environments.

II. Related Works

In the field of HAR, sensor data requires preprocessing followed by the application of recognition models to classify the activity classes of interest [11], [12]. Algorithms for identifying VDA from sensor data range from simple threshold-based models to sophisticated Machine Learning (ML) algorithms like deep learning. In the present work, we employ an ML-based classification model, and we review here several prior studies concerning the use of ML to classify VDA events.

One of the pioneering works on identifying VDA was performed in 1998 by Sagawa *et al.* [13], using accelerometer and barometer data to identify VDA. The study used a threshold-based model with a small training sample (83 minutes and 6 subjects). Since then, the combined effects of the digital revolution, the ubiquity of mobile devices, and the advances in sensors, Big Data, and ML have paved the way for the possibility of real-time recognition of activity classes of individuals evolving in complex environments.

Accelerometers are the most widely used sensors to track human activities [14]–[16]. Studies using an accelerometer as the stand-alone sensor show that the prevalent mode of tracked vertical mobility is stairs climbing [11], [17]–[23]. However, the classification accuracy of stairs climbing tends to be lower than other activity classes [24], [25]. This led researchers to include additional data sources to improve the accuracy of VDA detection, such as gyroscope, magnetometer, and barometer data [26]–[30], consider other modes of vertical mobility such as escalator and elevator rides. For instance, Liu *et al.* [26] added barometer data to a model that used accelerometer, gyroscope, and magnetometer sensor data, notably improving the classification accuracy from $\sim$ 80% to $\sim$ 90% (number of subjects: 10).

Additionally, some studies have acknowledged that accelerometers are effectively less robust than barometers for VDA classification. For example, Muralidharan *et al.* [31] compared the VDA recognition performance using an accelerometer versus a barometer and showed that their VDA classification performances were similar — with the barometer-based framework performing slightly better at nearly 100% accuracy (number of subjects: 2). However, the accuracy of the accelerometer-based framework dropped drastically when the mobile device was used to take calls or play games [31]. Similarly, Vanini *et al.* [32] showed that the classification performance of VDA was comparable for accelerometer-only and barometer-only study ($\sim$ 99%), but that the barometer was more energy-efficient and less dependent from the on-body position than accelerometers (number of subjects: 10).

Our study focuses on recognizing VDA as a general class of activity. It uses the following sensor data: (1) location data derived from Wi-Fi Access Points (APs), (2) magnitude of 3-axis accelerometer, and (3) barometric pressure data.

The magnitude of the 3-axis accelerometer signal is orientation independent [33], [34] and does not require complex data post-processing [11]. On the other hand, raw barometric sensor data can entail noise introduced by random sensor errors, limited sensor resolution, and high sampling frequency ($>$ 2 Hz). Filtering techniques like moving average filters [35]–[37], Finite Impulse Response (FIR) filters [38], and Infinite Impulse Response (IIR) filters [27], [38]–[40] are commonly used to alleviate the noise effects. Signal modeling such as sinusoidal fitting [41] and sigmoidal nonlinear fitting [42] is used to increase the precision in extracting elevation changes. In the present study, the spectral resolution of the sensor data collection is 0.06 Hz. Such a low sampling rate allows the system to side-step noise appearing at a high sampling rate that affects precise extraction of elevation changes. Therefore, the barometric pressure sensor data we consider here are not filtered, and other noise sources due to sensor resolution are used to quantify the uncertainty in the magnitude of the predicted VDA.

Barometric pressure data is usually converted to several common feature types such as: (1) statistical [43], [44], (2) spectral [43]–[45], (3) temporal [43], [44], and (4) wavelet-based features [46], [47]. The most commonly used fea-

tures are the rate of change of pressure (vertical velocity or slope) [26], [35], [43], [48], [49] and differential pressure (dp) [48], [50].

Vanini *et al.* [32] used barometric pressure data alone to recognize VDA using the features–rate of change of pressure and the standard deviation of differential pressure. They found that Long Short-Term Memory (LSTM) neural network framework produces a 99% accuracy compared to a decision tree approach (96%) and naïve Bayes classifiers (93%). However, their data collection was of short duration (30 minutes for each class) and conducted in limited environments. Muralidharan *et al.* [31] detected floor changes with an accuracy of 99% using the J48 decision tree model. Even though several factors that affect barometric pressure are considered, the data collected were of short duration (few minutes), conducted in limited structured environments, and lacked any entanglement with transportation modes. Liu *et al.* [26] classified vertical displacement activities from horizontal displacement activities (HDA) using inertial measurement units (IMU, including magnetometer) and barometer sensor data, with barometric sensor features derived from the standard deviation of pressure and rate of change of pressure. By training various classifiers such as Random Forest, J48 decision trees, Artificial Neural Networks (ANN), SVM, and Naïve Bayes, they obtained that Random Forest classifiers produced the highest accuracy of 92%. Also, in this study, each activity class was performed only for a few minutes and limited to ambulation [26].

The review of the literature on HAR [11], [12], [51]–[54] makes it clear that no classifier can be considered the best one universally, i.e. without considering the context in which it is used. As each data set comes with its own set of distinct characteristics, the classifier working best for a particular data set and activity type might not have the best performance for a specific problem or different circumstances (i.e., not generalizable) [11]. In this study, we have chosen two ensemble models (XGBoost and Random Forest), a neural network model (Multilayer perceptron) a Bayesian model (Naive-Bayes), and an instance-based nearest-neighbor model (k-Nearest Neighbor) to evaluate and compare the performances of each model on our data set (see Fig. 1).

It is common in the HAR literature to use short-duration training data collected in segments that contain only one or two activity classes and are performed in semi-natural or laboratory conditions, with limited variability in environments. However, real-life human activities occur in complex and unstructured environments, with a wide range of possible sequences, spanning heterogeneous activity classes with heterogeneous durations. Our study collocates itself in such a framework, as it uses a long-term (5 days) data set collected in a large-scale student population ($\sim 50,000$ students) during their regular weekdays. Hence, it requires a different approach than those reported in the literature.

Indeed, long-term monitoring of human activities requires a thorough understanding of all the factors affecting the sensor data in different static and dynamic environments. In particular, the factors that influence barometric pressure data are climate and weather, air velocity during motion, built environment, altitude, and sensor accuracy [9]. For a detailed review on the use of barometers to track human activity and the many factors that affect barometric pressure, we refer the reader to the recent review paper [9].

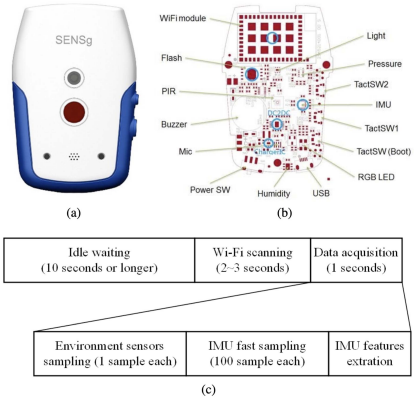


Fig. 2. NSE SENSg device details: (a) outside look, (b) internal structure [56], (c) working cycle. (Picture Courtesy: [57]).

III. Description of Data

A. National Science Experiment Data

The National Science Experiment (NSE) was designed and commissioned by the Singapore National Research Foundation (NRF) and the Singapore University of Technology and Design (SUTD), with other private and government bodies in Singapore [55]. The primary objective of this island-wide science experiment carried out by Singapore students—themed "Step Out for Science"—was to monitor and evaluate their carbon footprint, travel mobility patterns, amount of time spent indoors and outdoors, and more.

Almost $50,000$ students from 92 schools distributed nation-wide participated in the NSE in 2016 (Table I). Each student carried a wearable device called SENSg (see Fig. 2), which consisted of built-in environmental, motion sensors, and communication units. The devices were able to record and transmit the sensed data related to the Activities of Daily Living (ADL) of the students to a cloud server [56]. The data was recorded every $13 \sim 18$ seconds over weekdays, from Monday to Friday, for eight weeks in 2016. The device goes to sleep mode to reduce battery consumption when the processed IMU signal shows no user movement. The SENSg devices were handed over to the students on Mondays and collected back on Fridays. Hence, the full-time scale of daily ADL is only available on Tuesdays, Wednesdays, and Thursdays. Table II shows the cleaned NSE 2016 database after removing devices based on two criteria (a) data coverage for less than 6 hours per day (b) percentage of missing location data larger than 50% in a day.

Our study uses the accelerometer, barometer, and Wi-Fi scanners embedded in the SENSg devices to detect students' Vertical Displacement Activities (VDA) during their ADL. These sensor types are widely available in modern

TABLE I
NSE 2016 - DEVICE DELIVERY INFORMATION

Dates	No. schools	No. devices	Primary (6–12 y.o.)	Secondary (12–16 y.o.)	Pre-University (16–19 y.o.)
Apr 11–15	7	3,002	1,590	1,192	220
Apr 18–22	7	3,290	2,300	990	–
May 16–20	10	2,673	820	1,624	229
May 23–27	2	840	–	220	620
July 11–15	32	10,751	3,834	6,677	240
July 18–22	23	7,934	5,233	2,181	520
July 25–29	8	15,029	220	890	13,919
Aug 17–19	3	6,000	–	–	6,000
Total	92	49,519	13,997	13,774	21,748

TABLE II
CURATED NSE 2016 DATABASE

Dates	No. schools	No. IDs	Primary (6–12 y.o.)	Secondary (12–16 y.o.)	Pre-University (16–19 y.o.)
Apr 11–15	7	1,797	1,049	608	140
Apr 18–22	7	1,793	1,265	528	–
May 16–20	10	700	239	386	75
May 23–27	–	–	–	–	–
July 11–15	33	4,387	2,048	2,114	73
July 18–22	21	3,086	2,292	666	128
July 25–29	8	3,345	–	93	3,057
Aug 17–19	3	1,380	–	–	1,380
Total	89	16,581	6,939	4,395	4,853

smartphones, rendering them ideal for this particular HAR. The SENSg device comprises the IMU sensor MPU9250 from InvenSense and the barometer sensor BMP280 from Bosch Sensortech. The device also collected and stored up to a maximum of 20 Wi-Fi Access Points (AP) with the highest Receiver Signal Strength Indication (RSSI). We used Skyhook, a mobile location service from Boston, Massachusetts, that has geolocation of billions of Wi-Fi APs around the world, to convert the Wi-Fi APs to location coordinates [58]. The location coordinates (latitude and longitude) have a typical location accuracy of ± 100 meters. The location accuracy is increased by applying regression to the time-series of location data (see Appendix B). When the Wi-Fi APs are sparse or absent, location data is considered missing. Interpolation is then applied to the time-series data to predict these missing values (see Appendix A). The full sensor characteristics are shown in Table III [56]. The embedded barometer sensor is capable of detecting up to 1-meter changes in height, i.e., ± 12 Pa [56]. The raw values measured by the accelerometer along its three axes were processed on-board the SENSg, and only descriptive statistics of these raw data were recorded: (1) $\max(M_{\text{acc}})$—the maximum value of the accelerometer's signal magnitude (2) $\text{std}(M_{\text{acc}})$—the standard deviation of accelerometer's magnitude, both sampled at 100 Hz during the one-second data acquisition temporal window, which occurred in its turn with frequency ~ 0.0625 Hz. The SENSg device, along with its working cycle, is shown in Fig. 2.

B. Video-Annotated Data

As the measure of the barometric pressure behavior is influenced by many factors, it is vital to have at our disposal

TABLE III
SENSOR CHARACTERISTICS OF SENSG DEVICE

Sensor	Model	Range	Accuracy	Units	Poll frequency
Wi-Fi	SN8205	–	–	–	0.062 Hz
Barometer	BMP280	300 to 1,100 hPa	± 12 Pa	Pa	0.062 Hz
Accelerometer	MPU9250	± 2 g	± 80 mg	mg/LSB	100 Hz (for every 0.062 Hz)

TABLE IV
VIDEO-ANNOTATED DATA

Mode	Total time (hours)
Car	0.4
Bus	1.5
Train	0.7
Cycle	0.8
Walking	1.5
Idle	0.3
VDA	0.4
Total	5.6

a data set with the corresponding ground truth for validation purposes. Therefore, we collected approximately 6 hours of sensor data annotated using video recording to this aim. This data set was recorded across different modes of horizontal (walking, idle, train, bus, car, and cycle) and vertical (elevator, escalator, and stairs) activity (Table IV). Two researchers collected data on different days and times using a SENSg device that was hung using a lanyard similar to the one used by students in the NSE. The video was recorded using a Go Pro Hero 6 mounted on the chest. We will use this data set to validate manual labeling methods described in Sec. IV-A.

IV. METHODOLOGY

As the NSE was conceived as a large-scale data collection with relatively high temporal resolution and for long durations, sensor data collection was optimized to save battery life and data bandwidth. Unfortunately, this compromised sensor resolution, sampling rate, and type of sensor data collected, which inevitably makes the VDA identification more challenging. This section thus details the machine learning framework that we developed to identify and extract VDAs from continuous temporal segments of the NSE data (Fig. 1). Section IV-A explains the manual labeling techniques, and sections IV-B, IV-C, IV-D, and IV-E encompass the machine learning framework. We refer to the Appendices for details related to the pre-processing of the location data (App. A and B), while App. C describes the classification model parameters, and App. D deals with model tuning.

A. Manual Labeling

Accurate annotation of sensor data is highly manpower intensive [59]. One solution is to manually label a small subset of the data based on expertise and validate the classification framework. We were indeed able to perform such a manual

labeling in a subset of the NSE database (see Sec. IV-D), using our general understanding of the factors affecting barometric pressure [9] and leveraging the data annotated with the help of the ground truth video (Sec III-B). In this manual labeling, we labeled each data point as either VDA or HDA, based on the unique characteristic profile of vertical transportation. As noted in [9], pressure changes caused by factors other than vertical transportation are mostly long-term variations (e.g., diurnal pressure cycle) or brief and intense transient spikes (e.g., Indoor-to-outdoor transition).

First, and based on our pilot experiments illustrated in [9], we distinguish VDA from HDA in our manual labeling by monitoring the following characteristics: (1) a low horizontal travel velocity, (2) body movements picked up by the accelerometer, and (3) an increase or decrease in barometric pressure. We note that high magnitude pressure changes that can be mistaken for VDA occur during transportation modes [9], but these transportation modes can be identified by tracking the location data trend, which should then reveal a high horizontal velocity.

For validation of the procedure, the classifier will be trained on the manually labeled data set and then applied to the data set described in Sec.III-B, for which the ground truth is available.

B. Data Segmentation

We segment the time-series sensor data into regions of significant and minor pressure changes to allow the classifier to focus on the main characteristic of VDA—i.e., the pressure-altitude relation. First, each data point is considered to a have significant pressure change based on a cut-off value ($dp_{i,\text{cut-off}}$), determined by considering three points: (1) we want to discard small pressure jumps that could correspond to other factors that yield pressure changes of similar magnitudes, such as slopes and indoor-outdoor transitions [9], (2) we want to consider pressure changes corresponding to at least 50% of the minimum vertical displacement of a single floor, and (3) the changes need to be consistent with the sensor resolution ($\pm$12 Pa).

The change in pressure (dp_i) for each data point is calculated from the difference of the time-series data $dp_i = P_i - P_{i+1}$, where P_i is the pressure datum at instant t_i. The time interval $dt_i = abs(t_i - t_{i+1})$ associated with dp_i should be less than a cut-off value $dt_{i,\text{cut-off}}$ to consider that data can be missing in times of inactivity, and the resulting variation in pressure values for large values of dt_i might then be due to the diurnal pressure cycle. The consecutive significant pressure changes in the same direction (positive change or negative pressure drop) are then grouped to form a segment, i.e. $S_{P_{i,n+1}} = \{P_i, P_{i+1}, \ldots, P_n, P_{n+1}\}$ for a pressure change sequence of $S_{dp_{i,n+1}} = \{P_i - P_{i+1}, P_{i+1} - P_{i+2}, \ldots, P_n - P_{n+1}\}$. Other features and sensor data are grouped using the same groups of indices as for the segmented pressure sequences.

In the case of manually labeled time-series data set, we first label each data point as either VDA or HDA (Sec. IV-A). Therefore, each segmented data might contain both data points labeled VDA and data points labeled HDA. We thus label each segment using a majority rule, and assigning a VDA label in case of a draw. A perfect data segmentation would allow each VDA segment to indicate a complete VDA event with no false positives or false negatives. The choices of $dt_{i,\text{cut-off}}$ and $dp_{i,\text{cut-off}}$ ultimately determine the performance of this method. Hence, we compute the F_1 score (a.k.a. F-measure of balanced F-score) of capturing a complete VDA event in each segment labeled as VDA in the manually labeled training data for a range of $dt_{i,\text{cut-off}}$ ([30, 50, 90, 120] sec) and $dp_{i,\text{cut-off}}$ ([20, 23, 25, 27, 30] Pa) values (Table V). The best F_1 score is obtained for values $dp_{i,\text{cut-off}}$ and $dt_{i,\text{cut-off}}$ equal to 25 Pa and 120 seconds, respectively. We thus perform the data segmentation with these parameter values, and the final classification described below will be performed on these segmented data.

TABLE V
SELECTION OF DATA SEGMENTATION PARAMETERS BASED ON F_1 SCORE

$dt_{i,\text{cut-off}}$	F_1 score (%)				
120	97.68	98.09	**98.15**	96.98	96.98
90	97.65	98.06	98.12	97.94	96.95
60	97.65	97.99	98.06	97.87	96.88
30	97.14	97.39	97.45	97.23	96.22
$dp_{i,\text{cut-off}}$	20	23	25	27	30

C. Feature Engineering

Model explainability is a growing focus in Machine Learning. Therefore, it is natural to start from features based on domain-specific knowledge to improve explainability. We use our data exploration and our understanding of the sensor data and the target event to be recognized to design several domain-specific features. Specifically, we compute the following features from the accelerometer, barometer, and Wi-Fi localization data.

1) Rate of Pressure Change dp/dt: It accounts for the pace of the VDA. This distinguishes the elevation change based activities from phenomena that unravel over slow temporal scales such as sensor drift and diurnal pressure cycle.

2) Modified Zero-Crossing Rate $\widetilde{zcr}$: The zero-crossing rate zcr is a temporal feature that counts the number of sign changes for a given signal during a particular time window. Here, we modified this feature to count sign changes only if the corresponding magnitude difference in pressure is $\geq$ 20 Pa. This conditioned $\widetilde{zcr}$ can indeed identify the pressure spikes due to factors such as weather and climate, built environment, air velocity during motion, or sensor accuracy, and distinguish them from one-directional pressure changes that occur during elevation changes.

3) Horizontal Travel Velocity dx/dt: As the location data are recorded latitude and longitude, we use the haversine formula (see Eq (1)) to calculate the great circle distance x between two locations. It is based on the assumption that the Earth is approximately spherical, a valid assumption for small distances such as those measured in the NSE data. The horizontal travel velocity dx/dt plays a crucial role in differentiating significant pressure changes of VDA from transportation-based activities [9]. Specifically, the great-circle distance x between

TABLE VI
PROPERTIES OF THE TRAINING-TEST SPLIT DATA SETS

data set	Ground truth	Size in duration	Number of time-series data points	Number of segmented data	Number of devices	Number of individuals	Number of schools	Pri.	Sec.	Pre-U.
Training (80%)	Manual	64 (days)	207,380	7,132	64	64	64	21	20	23
Test (20%)	Manual	17 (days)	55,092	2,166	17	17	17	8	2	7
Video annotated data	Video recording	6 (hours)	1,262	305	5	2	–	–	–	–

location coordinates (φ_1, λ_1) and (φ_2, λ_2) with φ the latitude, λ the longitude, and r the radius of earth, is given by:

$$x = 2r \arcsin\left(\sqrt{\sin^2\left[\frac{\varphi_2-\varphi_1}{2}\right] + \cos\varphi_1 \cos\varphi_2 \sin^2\left[\frac{\lambda_2-\lambda_1}{2}\right]}\right). \tag{1}$$

4) Statistical Features of Immediate Neighborhood in Time-Series Data N_i: By definition, a VDA event is always preceded and followed by an HDA event. However, during vertical mobility modes like the elevator (and sometimes during escalator rides), a person is potentially standing with no significant body movement. Likewise, during transportation modes such as car, bus, or train travel, the vehicle stops intermittently, leading to low horizontal travel velocity regions. Hence, calculating statistical features over an immediate neighborhood of each data point can entail a sequence of events during and around an activity of interest. More precisely, for each data point i, we compute on the time window $N = \{i-2, i-1, i, i+1, i+2\}$ (of approximate width ~ 80 seconds) and for both the horizontal travel velocity dx/dt and the accelerometer data ($\max(M_{\text{acc}})$ and $\text{std}(M_{\text{acc}})$) the ten following statistics: Minimum, Maximum, Average, Median, Mode, RMS (Root Mean Square), MAD (Median Average Deviation), Standard deviation, Variance, and IQR (Inter Quartile Range).

5) Statistical Features of Segmented Data S_i: Many of the statistical features considered for the immediate neighborhood of time-series data are not suitable for segmented data, as the length of the sequences is typically very small for most VDA events—$2 \sim 4$ or even less ($1 \sim 2$) for transportation modes during which the sign of slope dp/dt changes very often. Hence, we compute only each data segment's mean, median, and mode. This is done both for the original sensor data and for the statistical features of immediate neighborhood in the time series, the modified zero-crossing rate $\widetilde{zcr}$, the horizontal travel velocity dx/dt, and the accelerometer data ($\max(M_{\text{acc}})$, and $\text{std}(M_{\text{acc}})$).

In total, we compute 95 features of the segmented data (Sec IV-B). We refer to this set of features as Feature set-I.

D. Training-Validation-Test Data

The NSE 2016 data set is very diverse in terms of the number of students, unique devices, and the demography of individual participants. Hence, each train-validation-test data set should reflect this diversity. To ensure this, we select the data collected during one day by 81 students from 81 different schools, with an appropriate balance of school types (primary, secondary, and pre-university) and weekdays (limited to Tuesday, Wednesday, and Thursday) to form a representative sample (see Table VI. We manually label this data sample according to the steps described in Section IV. This sample data set is then randomly divided into training (80%) and test set (20%). The data collected by any single device is assigned either to training or testing as a whole (data collected by a single student cannot be split between training or testing). The classifier model is trained on the training set, and a 5-fold cross-validation tunes the model's hyperparameters. Once the best model parameters are identified, it is then tested against the test set for the final performance evaluation.

In addition, to validate the manual labeling, we use the trained classifier model on the video-annotated data described in Sec.III-B.

E. Classification Models

We have selected five commonly used classifiers in HAR [11]: (1) Extreme Gradient Boosting (XGBoost or XGB), (2) Random Forest (RF), (3) Naive-Bayes model (NB), (4) k-Nearest Neighbors (kNN), and (5) Multilayer Perceptron (MLP). Both XGBoost and Random Forest are decision-tree-based ensemble learning algorithms. The XGBoost algorithm is based on the boosting method that adds weak learners sequentially to reduce the model's loss function. In contrast, the Random Forest model is based on the bagging method that adds weak learners in parallel and uses the majority voting model to make final predictions. On the other hand, the Naive-Bayes model is a probabilistic learning algorithm based on Bayes' theorem that assumes strong independence between the features. Furthermore, the k-Nearest neighbor model is non-parametric and uses distance-based measures to find the k-nearest samples and a majority voting model to assign a class. Finally, a Multilayer perceptron is an artificial neural network with a single input and output layer, and at least one hidden layer. Due to its multiple (hidden) layers and its nonlinear activation, MPL is known to be suitable for data that is not linearly separable.

The hyperparameters of these classifiers are tuned through a grid search using a 5-fold cross-validation on feature set-I. A more detailed description of the model parameters and model tuning can be found in App. C and D respectively.

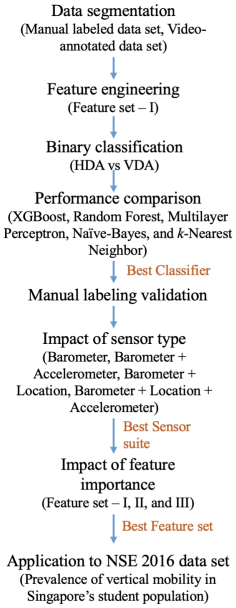


Fig. 3. Flowchart of Sec. V: Results and Discussion.

V. Results and Discussion

The central objective of our study is to obtain the magnitude of vertical displacements of individuals during their daily activities as an indication of vertical movement in a large-scale study of human mobility. By leveraging the relationship between barometric pressure and altitude, our specific aim is to develop pre-processing methods that alleviate the adverse influence of the other factors that affect barometric pressure to extract the instances of vertical motion accurately.

We globally consider the following procedure. First, a small representative sample of the cleaned NSE 2016 data (Sec IV-D) is manually labeled as described in Sec IV-A. We then segment the whole cleaned NSE 2016 data set from Table II and the manually labeled data and video annotated data according to Sec IV-B. The classifier model is then trained on the segmented manually labeled data to achieve a binary classification between Horizontal Displacement Activity (HDA) or Vertical Displacement Activity (VDA). Next, the trained classifier is applied to the video annotated data to validate manual labeling and finally applied to the cleaned NSE 2016 data set (see Fig. 3).

A. Performance Comparison of Classification Models

The four classifiers described in sec. IV-E are trained using feature set-I with the respective hyperparameters tuned from the 5-fold cross-validation with grid search (see Sec. IV-E). Their respective classification performances on the test data set are reported in Table. VIII. The XGBoost model provides the highest overall accuracy (98 %) and F_1-score for classifying VDA (93 %), closely followed by the Random Forest (overall accuracy of 97 % and F_1-score for classifying VDA at 89 %) and the Multilayer perceptron model (overall accuracy of 97 % and F_1-score for classifying VDA at 88 %). The Naive-Bayes model and k-Nearest Neighbor model perform poorly with an F_1-score for classifying VDA standing at 52 % and 17 % respectively. The ensemble learning methods clearly show a superior performance than Naive-Bayes and kNN models, similar to the results reported by Liu *et al.* [26] in classifying VDA from HDA. However, the neural-based MLP also produces similar overall performance as the ensemble learning models while falling slightly short in VDA classification performance. Therefore, we select XGBoost as the classifier of choice for this analysis.

It might be worth adding that future practitioners having access to much larger data sets—possibly with additional sensory features—may consider recent developments in Deep Learning [9], [51], [54], [78] to not only extract VDA but also to discriminate between all VDA types (e.g., escalator riding, staircase climbing, elevator riding).

As shown in Tab. VII, this study considers all the factors that influence barometric pressure with the data set that consisted of both general activity classes—ambulation and transportation—that a user comes across in a natural daily free-living environment and recorded over a long period ($1 \sim 3$ days). As a result, our results show high accuracy in detecting VDA (and HDA) similar to the results reported in the literature while greatly expanding the scope of the identification using the developed framework.

B. Validation of Manual Labeling

The trained XGBoost classifier model is applied to the data set with ground truth annotated from a video recording to ensure the validity of our manual labeling strategies. The results of the classifier performance is shown in Table IX. Although the classifier has 100% precision (i.e., it does not capture false positives), it has a recall of 80% (i.e., it only captures 80% of the actual VDA events), thereby yielding an F1-score of 89%.

The recall performance is relatively low due to a high proportion of instances such that $dp/dt < 1.9$ Pa/sec in the video-annotated data set (16.9% of all points marked as VDA) compared to the test data set (5.5%), for example. This issue is also responsible for some false negatives in the predicted data, where 77% of the predicted false negatives in video-annotated data set have $dp/dt < 1.9$ Pa/sec (see Fig. 4). This is because many of the data points with small pressure jump dp are not labeled as VDA in the manually labeled data set. Indeed, many instances with similar magnitudes for dp cannot be ruled out with high confidence, given the known factors influencing barometric pressure [9].

C. Impact of Sensor Type and Feature Importance

Three sensors are employed for this study: Barometer, tri-axis accelerometer, and Wi-Fi scanner (location data). To understand the impact of these sensor types on the

TABLE VII

CATEGORIZED LITERATURE RELATED TO STUDIES THAT HAVE IDENTIFIED VDA AS PART OF THE ACTIVITY CLASS USING BAROMETER AND ADDITIONAL SENSORS. SENSORS: BAROMETER (BARO), ACCELEROMETER (ACC), MAGNETOMETER (MAG), AND GYROSCOPE (GYRO). FACTORS CONSIDERED: 1. ALTITUDE, 2. CLIMATE AND WEATHER, 3. BUILT ENVIRONMENT, 4. AIR VELOCITY DUE TO MOTION, AND 5. SENSOR ACCURACY. ACTIVITY CLASS: AMBULATION (A) AND TRANSPORTATION (T)

Ref.	Sensors	Factors Considered	Activity Class(es)	Location	Time Period	Best ML model	Overall Accuracy (%)
This work	Baro, Acc, Wifi	1, 2, 3, 4, 5	A & T	Indoor and outdoor	Short and Long	XGBoost	98
[60]	Baro	1, 3, 4	T	Outdoor	Short	Threshold	84
[38]	Baro	1	A	Indoor	Short	Decision Tree	95
[48]	Baro	1, 2, 4, 5	A & T	Indoor and outdoor	Short	Threshold	93
[50]	Baro	1, 2, 3	A	Indoor and outdoor	Short and long	Naive-Bayes	99
[61]	Baro	1, 2, 3, 5	A	Indoor	Short	Threshold	99
[62]	Baro	1, 2, 5	A	Indoor	Long	Hierarchical clustering	98
[35]	Baro	1, 3, 5	A	Indoor	Short	Threshold	95
[31]	Baro	1, 2, 3, 5	A	Indoor	Short and long	J48 Decision Tree	100
[32]	Baro	1, 2, 3, 5	A & T	Indoor and outdoor	Short	Recurrent Neural Network	99
[63]	Baro	1, 2, 3, 4, 5	T	Outdoor	Short and long	DTW	80
[26]	Baro, Acc, Mag, Gyro	1, 2, 3, 5	A	Indoor	Short	Random Forest	92
[64]	Baro, Acc, Gyro	1	A	Indoor	Short	J48 Decision tree	80
[65]	Baro, Acc, Gyro, Mag, WiFi, Microphone	1	A	Indoor and outdoor	Short	Random Forest	85
[66]	Baro, WiFi	1, 2, 5	A	Indoor	Short and long	Monte Carlo Bayesian	97
[67]	Baro, WiFi	1, 2, 5	A	Indoor	Short	Bayesian	96
[27]	Baro, Acc, Gyro	1	A	Indoor	Short	KNN	97
[68]	Baro, Acc	1	A	Indoor	Short	Probabilistic Neural Network and Adjustable Fuzzy Clustering	91
[69]	Baro, Acc	1	A	Indoor	Short	SVM	98
[47]	Baro, Acc, Mag, Gyro	1, 3, 5	A	Indoor and outdoor	Short	SVM and KNN	96
[46]	Baro, Acc	1	A	Indoor	Short	Continuous Wavelet Transform	96
[39]	Baro, Acc	1	A & T	Indoor	Short	Threshold	96
[70]	Baro, Acc	1	A	Indoor and outdoor	Short and long	Decision Tree	97
[42]	Baro, Acc, Gyro, Mag, Foot pressure	1	A	Indoor	Short	Decision Tree	99
[71]	Baro, Acc	1, 5	A + cycling	Indoor and outdoor	Short	Decision Tree	98
[72]	Baro, Acc	1, 2, 3	A	Indoor	Short	Bayesian Network model	99
[73]	Baro, Acc, Gyro	1	A	Indoor	Short	Finite State Machine	NA
[74]	Baro, GPS	1, 2, 3, 4	T	Outdoor	Long	SVM	99
[44]	Baro, Acc	1	A	–	Short	Decision Tree	94
[13]	Baro, Acc	1, 2, 5	A	Indoor	Short	Kalman Filter	95
[75]	Baro, Acc, Mag	1, 3, 5	A	Indoor	Short	KNN and Random Forest	98
[41]	Baro, Acc, Gyro	1, 2	A	Indoor	Short	Decision tree	98
[76]	Baro, WiFi, Bluetooth	1, 2, 5	A	Indoor	Short	Threshold	90
[77]	Baro, Acc, Gyro, Mag, WiFi	1	A	Indoor	Short	Extended Kalman Filter	97
[43]	Acc, Gyro, Mag	1	A	Indoor	Short	Decision tree	90

classification of VDA, we conduct an ablation study. Since barometric pressure data is used in the pre-processing step (data segmentation) and is vital to extracting vertical displacements, features derived from barometer data are not removed

TABLE VIII
COMPARISON OF CLASSIFIER PERFORMANCE ON THE TEST DATA SET WITH FEATURE SET-I

Model	Overall accuracy (%)	Class	Precision (%)	Recall (%)	F_1-score (%)	Support
XGB	98	HDA	99	99	99	1899
		VDA	93	92	93	277
RF	97	HDA	98	99	99	1899
		VDA	96	83	89	277
MLP	97	HDA	98	98	98	1899
		VDA	89	88	88	277
NB	80	HDA	97	80	88	1899
		VDA	38	83	52	277
kNN	87	HDA	88	99	93	1899
		VDA	52	10	17	277

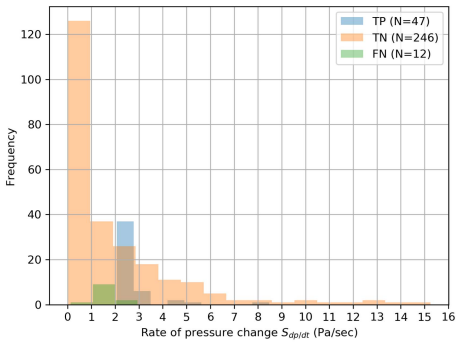


Fig. 4. Distribution of the rate of pressure changes in predicted data in video-annotated data set when using feature set-I.

TABLE IX
CLASSIFICATION RESULTS OF VIDEO-ANNOTATED DATA SET WITH FEATURE SET-I

(a) **Performance metrics**

Class	Precision (%)	Recall (%)	F1-score (%)	Support
HDA	95	100	97	246
VDA	100	80	89	59
weighted average	96	96	96	395

(b) **Confusion matrix**

Class	HDA	VDA	Support
HDA	246	0	246
VDA	12	47	59

in this ablation study. Table X reports the results of the classification performance with features from the following sensor(s): (a) Barometer only, (b) Barometer and tri-axis accelerometer, and (c) Barometer and Wi-Fi scanner (location data). The F_1-score for classifying VDA is 57% (overall accuracy of 89%) with barometer features only. It is significantly improved by adding the tri-axis accelerometer features (overall accuracy of 96% and F_1-score for classifying VDA at 83%). Even

TABLE X
IMPACT OF SENSOR TYPE ON THE CLASSIFIER PERFORMANCE WITH SELECTED FEATURES OF FEATURE SET-I. BARO - BAROMETER, ACC - ACCELEROMETER, AND LOC - LOCATION DATA DERIVED FROM WI-FI SCANNER

Sensor	Overall accuracy (%)	Class	Precision (%)	Recall (%)	F_1-score (%)	Support
Baro.	89	HDA	94	94	94	1899
		VDA	58	56	57	277
Baro + Acc	96	HDA	98	97	98	1899
		VDA	83	84	83	277
Baro + Loc	97	HDA	98	99	99	1899
		VDA	92	88	90	277
Baro + Loc + Acc	98	HDA	99	99	99	1899
		VDA	93	92	93	277

better performance is achieved with the combination of location and barometer features with 97% overall accuracy and 90% F_1-score for classifying VDA. The addition of the accelerometer data to barometer and location data thus only improves the F_1-score by 3%. As noted in Sec. II, previous studies [31], [32] have obtained similar accuracy with accelerometer only or barometer only features. However, their training data was collected in a limited environment (e.g., ambulation only or transportation only activities and short time period). For a data set such as the one considered here, which was collected in a long term free-living environment, it thus seems that barometer only features are not enough to classify VDA with high accuracy; however, adding the data of even one additional sensor data allows excellent performances even in this case.

The XGBoost model (see App. C for a description of the model and its hyperparameters) has an embedded feature ranking method that quantifies the importance of each feature to build the predictive model. The most relevant parameter to quantify relative feature importance is the total gain that measures the improvement in accuracy brought on by the feature for each tree in the model. Based on the total gain, the feature importance of all 95 features is calculated from the tuned model. To understand the impact of using reduced feature sets, we use the top 10 (out of 95) and the top 5 (out of 95) of the important features, denoted as the feature set-II (features 1–10 in Table XI) and feature set-III (features 1–5 in Table XI) respectively. We re-tune the hyperparameters and re-train the classifier model using the same procedure as described in App. D. The reduced feature set size of 10 and 5 are arbitrarily chosen, but it is supported by the fact that the total gain drops drastically after the most important feature ($S_{dp/dt}$) as shown in Table XI.

The classification performance of the XGBoost model with feature sets I, II, and III are shown in Table XII. The overall classification accuracy of the model for all feature sets on the test set stands at 98%, and the F_1-score for classifying VDA is found to be slightly lower at 93% for feature set-I and 92% for feature set-II and III. As the number of features is reduced

TABLE XI
FEATURE IMPORTANCE OBTAINED FROM THE TUNED XGBOOST CLASSIFIER MODEL. TOP 10 RESULTS ARE SHOWN. *S* STANDS FOR SEGMENTED DATA

No	**Feature**	**Total gain**
1	$S_{dp/dt}$	4460
2	$S_{\min(N_{\text{RMS(std}(M_{acc}))})}$	2614
3	$S_{\min(N_{\text{median}(dx/dt)})}$	988
4	$S_{\min(\widetilde{zcr})}$	421
5	$S_{dx/dt}$	265
6	$S_{\min(N_{\max(M_{acc})})}$	184
7	$S_{\min(N_{\max(dx/dt)})}$	184
8	$S_{\min(N_{\text{avg}(dx/dt)})}$	141
9	$S_{\max(N_{\max(dx/dt)})}$	124
10	$S_{\min(N_{\min(dx/dt)})}$	114

TABLE XII
CLASSIFICATION PERFORMANCE OF THE XGBOOST MODEL

(a) **Performance metrics**

Feature set	Overall accuracy (%)	Class	Precision (%)	Recall (%)	F_1-score (%)	Support
Feature set-I	98	HDA	99	99	99	1899
		VDA	93	92	93	277
Feature set-II	98	HDA	99	99	99	1899
		VDA	93	90	92	277
Feature set-III	98	HDA	98	99	99	1899
		VDA	94	90	92	277

(b) **Confusion matrix**

Feature set	Class	HDA	VDA	Support
Feature set-I	HDA	1870	19	1899
	VDA	22	255	277
Feature set-II	HDA	1870	19	1899
	VDA	27	250	277
Feature set-III	HDA	1874	15	1899
	VDA	29	248	277

from 95 to 10 and ultimately down to 5, there is a slight increase in precision with a complementary decrease in recall. The confusion matrix in Table XIIb shows no significant drop in performance when reducing the feature set size. We use the XGBoost model results with feature set-II in this section's further analysis to balance between performance and the number of features.

D. VDA Classification Performance

Table XIIb shows the confusion matrix—i.e., predicted class distributions and corresponding Type-I (False Positives) and Type-II errors (False Negatives). When closely inspecting the feature space in the test data set, one finds that the Type-II error generally occurs when the rate of pressure changes $S_{dp/dt}$ is small, typically below 1.9 Pa/sec for 65% of the false negatives. This observation may have two possible origins. This is either due to the small magnitude of pressure jumps dp (associated with low recording frequency that splits a single VDA event across multiple time intervals). Alternatively, this could also be due to larger time intervals dt (due to irregular recording frequency of 0.076 $\sim$ 0.016 Hz). The low and irregular sampling rate is the critical limiting factor in both cases. In addition, in some cases, a vertical mobility event follows or precedes a horizontal transportation mode—e.g., above-the-ground or underground train travel—, which may lead to misclassification due to the reliance on Wi-Fi localization to calculate the travel velocity. This is the case, for instance, when the APs are sparse.

E. VDA Recognition Limitations

The magnitude of the altitude change in a given VDA event is derived from the barometric pressure change of the segmented data S_{dp} using Eq (2). The accuracy of this magnitude is limited by the barometric sensor resolution, which is $\pm$12 Pa for our SenSg device (see Table III). Some of the VDA events with lower values of $S_{dp/dt}$ are not appropriately classified, as stated in Sec. V-D. For a typical sampling frequency of 0.062 Hz, this corresponds to an altitude change of 2.5 meters (30 Pa). Hence, it is reasonable to assume that vertical displacements smaller than this value are not properly captured, and thus the vertical moves can only be accurate with a vertical resolution of $\sim$ 2.5 meters.

The classification performance when recognizing VDAs in this study is thus limited by the type of sensor data, sensor resolution, and sampling frequency. For example, location data from GPS with an accuracy of $\pm$10 meters would outperform the often inaccurate Wi-Fi localization, which is only accurate within $\pm$200 meters. Similarly, a higher sampling rate of barometric pressure to the tune of 1 Hz would be ideal compared to the lower sampling frequency of $\sim$ 0.062 Hz in this data set. Moreover, using an accurately annotated data set—e.g., employing video recording—can markedly improve the training performance.

VDA can be further sub-classified into different modes of vertical transportation like elevator, escalator, or stairs. This would, however, require large amounts of video-annotated training data set with high temporal and spatial resolution. This is because the rate of change of vertical displacement—a key metric in distinguishing between these modes [9]—needs to be sampled at a reasonably high resolution, especially to differentiate between stairs climbing and escalator riding. Furthermore, the classification between stairs/escalators and elevators will only be accurate for floor jumps larger than 2 $\sim$ 3 if the sampling frequency is low, such as that found in our data set. For these reasons, we do not consider VDA sub-classification in this study. However, using sensors with a higher sampling rate, such sub-classification of VDA should be attainable, as shown in previous works [26], [31], [32], [75].

F. Prevalence of Vertical Mobility in Singapore's Student Population

We have applied the trained XGBoost model with feature set-II to the curated NSE data described in Table II. The data

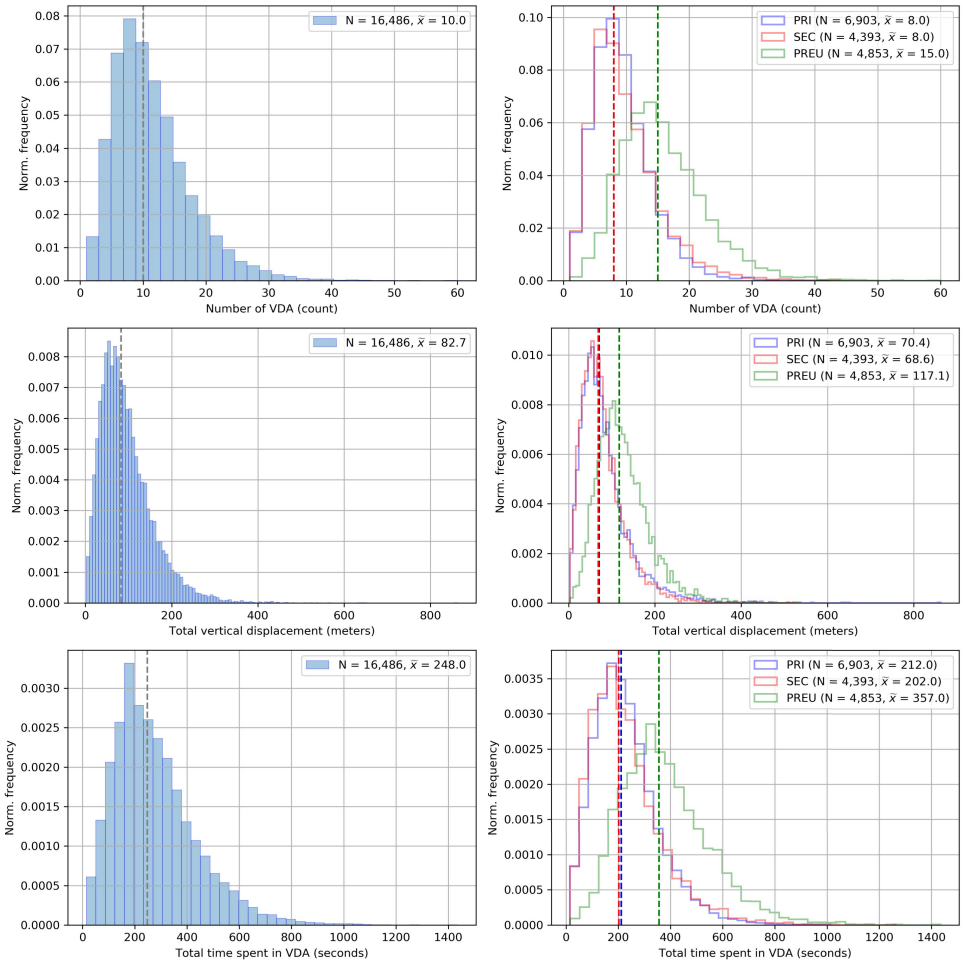


Fig. 5. Aggregated distribution of VDA for a population of $N = 16{,}486$ individuals. The aggregation period is set to 1 day. N is the number of subjects and $\tilde{x}$ denotes the median value (dashed lines). Top row: number of VDA events per subject; Middle row: total cumulated vertical displacement (in meters) per subject; bottom row: total time spent in VDA (in seconds) per subject; left column: total population; right column: subjects grouped by school types.

was collected for $N = 16,486$ students from 89 schools aged 6–19 whose residences and schools were spread throughout the island-state of Singapore. For each subject, the data contains at least 6 hours of coverage during a day. The results presented in this section are aggregated over a single day for each participant—the selected day corresponds to the one with the largest data points collected during their week of carrying the SENSg device as part of the NSE program. In addition, the predicted VDAs are post-processed to remove vertical displacements (less than 9% of total) that are accompanied by significant pressure fluctuations ($dp > 20$ Pa) that may have been caused by sensor errors (see [9]).

The predicted VDAs in the entire data set reaches 182, 841 events after postprocessing. Some statistics of VDA for each subject over a day are calculated and shown in Fig. 5, namely the number of VDAs, the cumulated vertical displacement, and the total time spent in the VDA mode. On a daily average, a subject was found to be engaged in 10 events of vertical mobility, traveled vertically $\sim$ 83 meters, and spent a total of 4 minutes per day in this mode. As an element of comparison, Americans are found to spend on average $\sim$ 65 minutes per day eating [79]. However, the distributions are rather heterogeneous, with individuals who traveled as much as 140 meters vertically in a single VDA event, moved

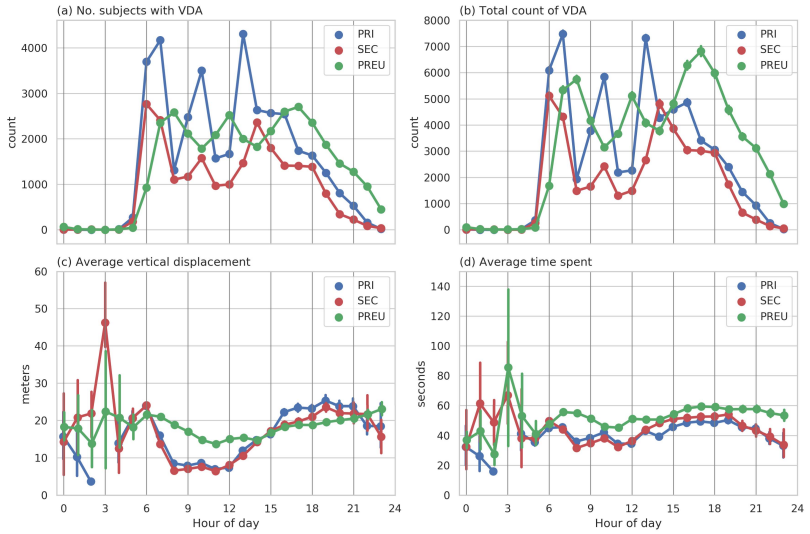


Fig. 6. Temporal statistics of VDA according to group types: (a) number of subjects with VDA (active subjects) by hours, (b) total VDA count by hour, (c) average vertical displacement per active subject by hour, and (d) hourly average time spent in VDA per active subject by hour. The errorbars in (c) and (d) indicate the confidence intervals estimated using a bootstrap approach ($n = 1000$).

up to 60 times in vertical direction and spent up to 24 minutes per day in VDA. The highest vertical displacement in our predicted data (140 meters) is nearly half the size of the highest building in Singapore—the Guoco tower stands at 290 meters.

Fig. 5 also shows the breakdown of the distributions of VDA for primary (PRI), secondary (SEC), and pre-university (PREU) students. Interestingly, the vertical mobility footprint of PREU students is markedly higher than that of PRI or SEC students, even if the descriptive statistics of single VDA events are very similar: the median values are 5.2 m for PRI students, 5.1 m for SEC students, 5.8 m for PREU students; the mean values are respectively 9.3 m (PRI), 8.3 m (SEC), 8.1 m (PREU) and the maximum values are 116 m (PRI), 120 m (SEC), 140 m (PREU). The fact that pre-university students are more active in terms of cumulated VDA can, in particular, be explained by the fact that they are more likely to travel through public transportation like trains that are either above or below the ground road level [80].

The timeline of VDA reveals a rich structure. In highly vertical cities, people experience significant waiting times during their vertical mobility due to shared public transportation such as elevators or slow pedestrian movements in escalators and stairs. In Singapore, these waiting times can be similar in magnitude to the waiting times between trains or buses ($2 \sim 5$ minutes). Businesses operating on appointments only often remind their customers to include the waiting time associated with elevator rides when planning their arrival. While these time scales can still be considered small, the ongoing vertical integration of multi-purpose buildings will underscore the importance of understanding congestion in vertical transportation and mobility. Understanding congestion per land-use type and time is essential to inform better transportation planning, and this point will also apply to vertical transportation.

Fig. 6 shows the hourly statistical distribution of VDA according to the three group types under investigation (PRI, SEC, and PREU). The group-based total number of subjects with VDA (i.e., active subjects, Fig. 6(a)) and total count of VDA (Fig. 6(b)) exhibit similar trends. They indicate that different daily rhythms of individuals occur between the three groups, i.e., the peaks in activity for PRI are at 7 a.m., 10 a.m., 1 p.m., and 4 p.m., SEC at 6 a.m., 10 a.m., 2 p.m., and 6 p.m., and PREU students at 8 a.m., 12 p.m., and 5 p.m. In addition, a large number of activities of PREU students is observed during evening and night times, which can readily be explained by the fact that some schools like the Institute of Technical Education (ITE), operate primarily in the evening and by the active after-school life of the late-teen population.

Due to the low number of active users between 12 a.m. and 5 a.m., the average vertical displacement (Fig. 6(c)) and average time spent in VDA (Fig. 6(d)) show considerable uncertainty within that period and become stable again after 5 a.m. In terms of per-subject average vertical displacement and time spent, the primary and secondary students show similar patterns, while pre-university students yield markedly different patterns. For instance, the primary and secondary students' average vertical displacement is high during the 5 a.m.–6 a.m. time window, drops to 9 meters between

8 a.m.–12 p.m., then peaks up again to $\geq$ 20 meters between 4 p.m.–9 p.m., and decreases after that. On the other hand, the pre-university students' average vertical displacement is about 20 meters between 6 a.m.–7 a.m., and 14 meters around 11 a.m., then slowly increases until 11 p.m. These differences in the morning could be explained by the style of classes between them. Primary and secondary classes are usually fixed, and students do not need to change classes between courses. In contrast, pre-university students will need to move to other classes for different courses.

VI. Conclusion

Over the past two decades, advancements in sensor technology and complexity science have enabled the dissection, with unprecedented accuracy, of the fine details of human mobility in urban areas from large-scale data. Mining these "Big Data" revealed the burstiness and relative predictability of human spatial movements [81], [82], as well as highly dense spatial areas in cities [83]. However, these studies, albeit illuminating, were mainly carried out on large groups of people living and moving about in mostly flat cities and were therefore limited to horizontal movements in two dimensions. The present study explored the application of identifying and monitoring VDA to highlight the prevalence of vertical mobility at city-scale and its potential to add to the understanding of human mobility in general.

This "Big Data" set consisted of 16,581 students between the age of 6 and 19 from 89 schools spread throughout the city-state of Singapore. The relatively long-term tracking (5 days) and high recording frequency (every $\sim$ 16 seconds) have provided high spatiotemporal resolution data sufficient to identify VDA in diverse conditions. The classification performance of the model is similar or closely aligned with the results reported in the literature (Sec. II)—with 98% overall accuracy and 92% F_1-score in classifying VDA—while significantly generalizing its applicability to almost all possible real-world conditions.

The recognition of VDA is instrumental to the effectiveness of a range of critical applications in indoor positioning and navigation [29], [30], [36], [66], estimation of energy expenditure, and health monitoring [16], [39], [47]. The present study markedly expands the scope of applications by rendering possible the performance of large-scale human experiments with the ultimate goal of shedding light on the human mobility patterns in dense urban areas, and their impact on urban planning. Furthermore, the VDA extraction methodology described in this paper can be used to perform urban data analytics. These results may have significant ramifications for the design of dense vertical cities. For instance, using this methodology, one should identify the hidden vertical barriers and possible facilitators to movement in the vertical city. Urban designers and planners would then have quantitative information about the hitherto unknown patterns of vertical human mobility. This new knowledge could prove pivotal to establishing novel design principles that can enhance mobility and, therefore, livability in the vertical urban world.

Our study is ultimately limited by the participants' age group and size, the sensor characteristics (primarily resolution and sampling rate), and the training data collection methods (video annotation versus manual labeling). Future studies can most likely achieve more significant results by improving one or more of these categories to obtain a more accurate and general detection of VDA.

Lastly, aggregated VDA data analysis by population groups/segments has revealed rich details about the anticipated differences in the daily activity profiles between late-teens and younger children. The uncovered patterns of vertical human mobility enabled us to accurately quantify, for the first time, the distance traveled and time spent in vertical transportation in a densely built urban environment. It has also revealed some unique patterns of activity related to vertical transportation that is present in many aspects of our human lives. There is no doubt that a more systematic analysis of the non-aggregated data would provide important new details and unique information about the dynamics of vertical mobility across several dimensions, including gender, age, and socioeconomic status.

Appendix

A. Localization & Interpolation

The location data is obtained using an API from a third-party company called Skyhook based on the available Wi-Fi APs. It requires a minimum number of APs to triangulate a given location. As the density of Wi-Fi APs varies from place to place, accurate localization may not always be achievable with insufficient data. This is especially pronounced in transportation modes—such as underground subway rides, by-pass roads, tunnels, and bridges.—where Wi-Fi APs are sparse or non-existent, even in highly dense cities. In our dataset, a per-person average of 8% data points are missing location information due to failed localization.

An interpolation method is used to predict the missing values to complete the Wi-Fi localization data. As the first step, successfully localized time-series are selected and a fraction of its data is removed randomly. Next, the Root Mean Squared Error (RMSE) value between original and randomly removed data is calculated for three distinct interpolation methods: (1) linear (2) cubic spline, and (3) piece-wise cubic spline to identify the best method. The linear interpolation consistently outperforms other methods under consideration with the lowest RMSE values for a fraction (10 $\sim$ 50%) of the data removed. For example, for fraction$>$ 20% of the data removed, RMSE measured in the degree of latitude/longitude is $\sim$ 0.0001 for linear interpolation while RMSE (degree)$\sim$ 0.005 for other methods.

B. Regression

The Wi-Fi localization data has a low spatial resolution of $\pm$400 meters. The travel velocity estimation based on this data reveals high local errors. A regression model is applied to find the best fit to reduce local fluctuations. To analyze the model performance, RMSE is used to estimate the residuals, and the R-squared value quantifies the proportion of variance explained by the model. The regression model performs poorly when applied to the entire time-series sequence. This is due to the different regions of variability in data. The time series

location data is segmented into a series of local (temporary) variable and global (long-term) variable sequences. The regression model is then applied with different window sizes or knot placements to each of these sequences, a higher number of knots for a global variable sequence and a lower number of knots for a local variable sequence.

A local variable sequence is defined as $S_{i,n} = \{i, i+1, \ldots, n\}$ where the location of index $-\, n$ is at distance $\leq$ distance_cut-off from index $-\, i$. A global variable sequence is created by combining several local variable sequences of length < 10 (~ 2.5 minutes) to reflect long term changes in location.

Two regression models: (1) piece-wise polynomial and (2) natural piece-wise cubic regression spline are considered as well as three smoothing models: (3) Savitzky-Golay smoothing, (4) LOESS model, (5) Exponential smoothing model. These five options are compared against different window sizes or knot placements (local variable $= [5, 15, 30]$, global variable$= [5, 15, 30]$) and distance_cut_off $= [0.1, 0.3, 0.5]$ km). The natural piece-wise Cubic regression spline and Savitzky-Golay show the lowest RMSE and highest R-Squared values for all knot placements/window sizes and distance_cut_off. Either of these methods will suitably reduce the local errors in location data. A low RMSE and high R-squared value can also point to over-fitted data; hence, additional considerations should support model selection. The optimal knot placements should be sparse for local variable data to reduce local fluctuations and denser for global variable data where location varies long-term. Here, Natural Piece-wise cubic regression spline method is selected with knot placements at (local variable, global variable)$= (15, 1)$ of the segmented data using distance_cut_off $= 0.3$ km. Essentially, the knot placements are designed to smooth the local variable data while leaving the global variable sequences intact.

C. Classification Model Parameters

1) XGBoost: Gradient boosting allows an ensemble of weak learners to build models that depend on the gradient descent algorithm to optimize an objective function. The objective function measures the model fitness of the training data and consists of training loss and regularization [84]. Training loss quantifies the model's predictive capability, and the regularization parameters help reduce overfitting by controlling the complexity of the model. The balance between these two terms is commonly known as the bias-variance trade-off [84]. XGBoost classifier model encompasses several parameters that are categorized as general, booster, and learning task parameters [85]. The tree booster parameters allows the specification of learning rate (eta), minimum loss reduction to make a split ($gamma$), maximum depth of a tree (max_depth), minimum sum of weights of all observations in a child (min_child_weight), fraction of observations to be sampled randomly ($subsample$), fraction of columns to be sampled randomly in each tree ($colsample\_bytree$), control of class imbalance ($scale\_pos\_weight$), L2 regularization term on weights (reg_alpha), and L1 regularization term on weights (reg_lambda) [84], [85]. Learning task parameters allows choosing an objective function ($objective$) for optimization against a specified evaluation metric ($eval\_metric$). The number of trees or estimators ($n\_estimators$) required depends on the learning rate eta. Optimizing the tree booster parameters based on the evaluation metric allows the objective function to look for values that avoid over-fitting. Model over-fitting can be further controlled by reducing the learning rate and increasing the number of estimators [85].

2) Random Forest: The Random Forest classifier is an ensemble learning algorithm that tries to create a range of uncorrelated trees by randomly selecting features for each tree and randomly (with replacement) assigning training data to each tree. Both of these properties of Random Forest helps control over-fitting [86], [87]. The parameter $n\_estimators$ gives the number of trees, and the maximum number of features considered for splitting a node is given by $max\_features$. The maximum number of levels in each tree is controlled by the parameter max_depth. The samples given to the tree are controlled by the following parameters: $bootstrap$, min_sample_split, $min\_weight\_fraction\_leaf$, and min_sample_leaf. The parameter $bootstrap$ can be set to $True$ to allow random sampling with replacement. The minimum number of data points assigned to a node before splitting is given by min_sample_split, and the parameter min_sample_leaf is used to control the minimum number of data points allowed in a leaf node. The sample weight can be adjusted using the parameter $min\_weight\_fraction\_leaf$. Finally, the class imbalance in a data set can be controlled by setting the parameter $class\_weight$ to $balanced$.

3) Naive-Bayes: The Naive-Bayes model is a probabilistic learning algorithm based on Bayes' theorem. This study uses the Gaussian Naive-Bayes algorithm that assumes Gaussian distribution for each class [87], [88]. Other Naive-Bayes models such as Multinomial and Bernoulli are defined for discrete data values and hence not suitable for our problem [87]. Gaussian Naive-Bayes model has only two parameters: $var\_smoothing$ and $priors$. The prior probability of the classes can be assigned through $priors$. Since there is no prior probability available for our classes, this parameter is unspecified in this work so that the priors can be learned from the data. The parameter $var\_smoothing$ is used to adjust the weight given to data points far from the mean distribution. This is the only parameter tuned for the Naive-Bayes model here.

4) k-Nearest Neighbors: k-Nearest Neighbors classifier is a class of Nearest Neighbors algorithm that identifies k-nearest training data points based on their similarity [87]. The number of neighbors k is given by the parameter $n\_neighbors$. The neighbors can be weighed uniformly or differently using the parameter $weight$. The distance metric used can be controlled by the parameter p, which denotes the power of the Minkowski distance ($p = 1$ denotes Euclidean and $p = 2$ denotes Manhattan). The speed of finding the nearest neighbors depends on the parameter $algorithm$ that can be set to brute-force or faster methods such as tree-based search algorithms. The tree-based search can be controlled by the parameter $leaf\_size$ that can be adjusted for faster construction and queries.

5) Multilayer Perceptron: The multilayer perceptron is a class of feed-forward artificial neural network models. This supervised learning model uses the backpropagation method that calculates the gradient of a loss function concerning the weights of the network to train the model [89]. It consists of a single input and single output layer and at least one hidden layer. The number of hidden layers and the number of neurons in each hidden layer can be selected using the parameter $hidden\_layer\_sizes$. The neurons can be activated or not using the activation function defined by the parameter $activation$. The weights of the network can be optimized by the solver (e.g., stochastic gradient descent) with the parameter $solver$. Similar to the XGBoost model, the L2 regularization is controlled by $alpha$. Another important parameter for MLP is the $learning\_rate$ that can be a given constant value, scaling value, or an adaptive parameter.

D. Classifier Model Tuning

The four classifier models follow an overall procedural structure for model tuning. The optimal model hyperparameters are selected by grid search and a 5-fold cross-validation using the metric – Area Under the Receiver Operating Characteristic Curve (ROC AUC). The grid search is initialized with hyperparameter values, which are fine tuned until no further change in the results is noticed. The 5-fold cross-validation is used to reduce the bias and efficiently uses the training set to understand the model's predictive power on new data during the tuning procedure. Due to the specific nature of the classifier models, some differences in the procedure exist. For example, all the hyperparameters for the Random Forest, Naive-Bayes, and k-Nearest Neighbors are tuned using a single grid search. In contrast, the hyperparameters for XGBoost are tuned in a series of grid searches. The following steps thus tune the XGBoost model:

XGBoost Model Tuning Steps: The optimal parameter values are calculated in a series of steps, where each step progresses by estimating the parameter under consideration based on the parameters calculated from the preceding step. The steps taken to tune the model are as follows: (1) Set a high learning rate and find the optimal number of estimators, (2) For the given learning rate and the number of estimators, find optimal max_depth and min_child_weight, (3) Find the optimal value of $gamma$ (4) Re-calibrate the optimal number of estimators, (5) find optimal values for the parameters $subsample$ and $colsample\_bytree$, (6) find the optimal values for the regularization parameters reg_alpha and reg_lambda, (7) Re-calibrate the optimal number of estimators, (8) Reduce the learning rate and find the optimal number of estimators [85].

The cross-entropy loss for binary classification is set as the objective function. The above steps are followed for an initial learning rate of 0.3 from (1) to (7). As the final step (8), the learning rate is reduced considerably (from 0.3 to 0.2 for feature set-I, to 0.1 for both feature set-II and feature set-III) to control over-fitting while keeping the F_1 score of VDA classification higher. The optimal number of estimators is then found to be 43, 125, and 72 for feature set-I, II, and III, respectively, from 5-fold cross-validation for the given learning rate. The class imbalance is acknowledged by setting the parameter $scale\_pos\_weight$ to 1 for faster convergence.

E. Pressure-Altitude Relation

The fundamental equation for fluids at rest dictates the relationship between atmospheric pressure and altitude [90]. The change in altitude is indeed given by

$$z_2 - z_1 = -\frac{p_2 - p_1}{\gamma}, \tag{2}$$

where $\gamma = \rho g$ is the specific weight of air with density $\rho = 1.225$ kg/m^3 and g is the acceleration due to gravity $g = 9.81$ m/s^2 at standard sea-level conditions. The fluid is assumed to be incompressible and in isothermal conditions. This pressure-altitude equation is valid for data collected for elevation less than 10 km from sea level [90], which is always the case in our study.

ACKNOWLEDGMENT

On-time production of the sensors was only possible due to strong support from Delta Electronics DRC, IABG, and Taoyuan Factory. Logistics for the distribution of thousands of sensors was facilitated by the Singapore Science Center, and the infrastructure to securely store the data was built and maintained by the A*STAR Institute for High Performance Computing.

The authors would like to thank Francisco Benita, Francesco Scandola, and Garvit Bansal for their assistance with SENSg devices and guidance on data collection. They would also like to thank the National Research Foundation (NRF) of Singapore for providing access to the National Science Experiment data.

REFERENCES

[1] Y. Zhou, *et al.*, "Understanding urban human mobility through crowdsensed data," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 52–59, Nov. 2018.
[2] E. Wilhelm, *et al.*, "Wearable environmental sensors and infrastructure for mobile large-scale urban deployment," *IEEE Sensors J.*, vol. 16, no. 22, pp. 8111–8123, Nov. 2016.
[3] Y. Zhou, B. P. L. Lau, C. Yuen, B. Tuncer, and E. Wilhelm, "Understanding urban human mobility through crowdsensed data," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 52–59, Nov. 2018.
[4] A. G. Yeh and B. Yuen, *Introduction: High-Rise Living in Asian Cities*. Dordrecht, The Netherlands: Springer, 2011, pp. 1–8.
[5] S. Gopalakrishnan, D. Wong, A. Manivannan, R. Bouffanais, and T. Schroepfer, "User-driven emergent patterns of space use in vertically integrated urban environments," in *Proc. 21st Int. Conf. Architectural Res. Centers Consortium (ARCC)*, Phoenix, AZ, USA, 2021, pp. 215–222. [Online]. Available: http://www.arcc-arch.org/wp-content/uploads/2021/08/ARCC-2021-Proceedings_Digital-Version_Web_PP.pdf
[6] R. Bouffanais and L. Sun Sun. (2019). *The Rise of Homo Verticalis*. Accessed: Feb. 22, 2022. [Online]. Available: https://blogs.scientificamerican.com/observations/the-rise-of-homo-verticalis/
[7] H. Barbosa *et al.*, "Human mobility: Models and applications," *Phys. Rep.*, vol. 734, pp. 1–74, Mar. 2018.
[8] S. Hasan, C. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *J. Statist. Phys.*, vol. 151, nos. 1–2, pp. 304–318, Apr. 2013.
[9] A. Manivannan, W. C. B. Chin, A. Barrat, and R. Bouffanais, "On the challenges and potential of using barometric sensors to track human activity," *Sensors*, vol. 20, no. 23, p. 6786, Nov. 2020.
[10] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition for health research," 2019, *arXiv:1910.03970*.

[11] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[12] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, p. 33, 2014.

[13] K. Sagawa, T. Ishihara, A. Ina, and H. Inooka, "Classification of human moving patterns using air pressure and acceleration," in *Proc. 24th Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 2, Aug./Sep. 1998, pp. 1214–1219.

[14] M. Janidarmian, A. Roshan Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors*, vol. 17, no. 3, p. 529, Mar. 2017.

[15] N. Twomey *et al.*, "A comprehensive study of activity recognition using accelerometers," *Informatics*, vol. 5, no. 2, p. 27, 2018.

[16] M. Saleh, M. Abbas, and R. B. Le Jeannes, "FallAllD: An open dataset of human falls and activities of daily living for classical and deep learning applications," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1849–1858, Jan. 2021.

[17] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[18] J. Mäntyjärvi, J. Himberg, and T. Seppänen, "Recognizing human motion with multiple acceleration sensors," in *Proc. IEEE Int. Conf. Syst., Man Cybern. e-Syst. e-Man Cybern. Cyberspace*, vol. 2, Oct. 2001, pp. 747–752.

[19] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," *Sensors*, vol. 19, no. 7, p. 1716, Apr. 2019.

[20] K. Kunze and P. Lukowicz, "Sensor placement variations in wearable activity recognition," *IEEE Pervasive Comput.*, vol. 13, no. 4, pp. 32–41, Oct. 2014.

[21] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.

[22] L. Cong, J. Tian, and H. Qin, "A practical floor localization algorithm based on multifeature motion mode recognition utilizing FM radio signals and inertial sensors," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8806–8819, Aug. 2020.

[23] S. Boim, G. Even-Tzur, and I. Klein, "Height difference determination using smartphones based accelerometers," *IEEE Sensors J.*, early access, Mar. 2, 2021, doi: 10.1109/JSEN.2021.3062346.

[24] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, and K. Aberer, "Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach," in *Proc. 16th Int. Symp. Wearable Comput.*, Jun. 2012, pp. 17–24.

[25] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, vol. 3001. Berlin, Germany: Springer, 2004, ch. 1, pp. 1–17.

[26] M. Liu, H. Li, Y. Wang, F. Li, and X. Chen, "Double-windows-based motion recognition in multi-floor buildings assisted by a built-in barometer," *Sensors*, vol. 18, no. 4, p. 1061, Apr. 2018.

[27] A. Moncada-Torres, K. Leuenberger, R. Gonzenbach, A. Luft, and R. Gassert, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," *Physiol. Meas.*, vol. 35, no. 7, p. 1245, 2014.

[28] A. El Halabi and H. Artail, "Integrating pressure and accelerometer sensing for improved activity recognition on smartphones," in *Proc. 3rd Int. Conf. Commun. Inf. Technol. (ICCIT)*, Jun. 2013, pp. 121–125.

[29] M. Ji, J. Liu, X. Xu, and Z. Lu, "The improved 3D pedestrian positioning system based on foot-mounted inertial sensor," *IEEE Sensors J.*, vol. 21, no. 22, pp. 25051–25060, Nov. 2021.

[30] Y. Zhao, J. Liang, Y. Cui, X. Sha, and W. J. Li, "Adaptive 3D position estimation of pedestrians by wearing one ankle sensor," *IEEE Sensors J.*, vol. 20, no. 19, pp. 11642–11651, Oct. 2020.

[31] K. Muralidharan, A. J. Khan, A. Misra, R. K. Balan, and S. Agarwal, "Barometric phone sensors: More hype than hope!" in *Proc. 15th Workshop Mobile Comput. Syst. Appl.*, 2014, p. 12.

[32] S. Vanini, F. Faraci, A. Ferrari, and S. Giordano, "Using barometric pressure data to recognize vertical displacement activities on smartphones," *Comput. Commun.*, vol. 87, pp. 37–48, Aug. 2016.

[33] S.-M. Lee, S. Min Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2017, pp. 131–134.

[34] I. Cleland, M. P. Donnelly, C. D. Nugent, J. Hallberg, M. Espinilla, and M. Garcia-Constantino, "Collection of a diverse, realistic and annotated dataset for wearable activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2018, pp. 555–560.

[35] C. Bollmeyer, T. Esemann, H. Gehring, and H. Hellbruck, "Precise indoor altitude estimation based on differential barometric sensing for wireless medical applications," in *Proc. IEEE Int. Conf. Body Sensor Netw.*, May 2013, pp. 1–6.

[36] G. Pipelidis, O. R. M. Rad, D. Iwaszczuk, C. Prehofer, and U. Hugentobler, "A novel approach for dynamic vertical indoor mapping through crowd-sourced smartphone sensor data," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2017, pp. 1–8.

[37] A. M. Sabatini and V. Genovese, "A sensor fusion method for tracking vertical velocity and height based on inertial and barometric altimeter measurements," *Sensors*, vol. 14, no. 8, pp. 13324–13347, Aug. 2014.

[38] B. Ghimire, C. Nickel, and J. Seitz, "Pedestrian motion state classification using pressure sensors," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–6.

[39] F. Bianchi, S. J. Redmond, M. R. Narayanan, S. Cerutti, and N. H. Lovell, "Barometric pressure and triaxial accelerometry-based falls event detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 6, pp. 619–627, Dec. 2010.

[40] M. Voleno, S. J. Redmond, S. Cerutti, and N. H. Lovell, "Energy expenditure estimation using triaxial accelerometry and barometric pressure measurement," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 5185–5188.

[41] F. Massé, R. R. Gonzenbach, A. Arami, A. Paraschiv-Ionescu, A. R. Luft, and K. Aminian, "Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients," *J. Neuroeng. Rehabil.*, vol. 12, no. 1, p. 72, Dec. 2015.

[42] C. Moufawad el Achkar, C. Lenoble-Hoskovec, A. Paraschiv-Ionescu, K. Major, C. Büla, and K. Aminian, "Instrumented shoes for activity classification in the elderly," *Gait Posture*, vol. 44, pp. 12–17, Feb. 2016.

[43] M. Elhoushi, J. Georgy, A. Wahdan, M. Korenberg, and A. Noureldin, "Using portable device sensors to recognize height changing modes of motion," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2014, pp. 477–481.

[44] C. Figueira, R. Matias, and H. Gamboa, "Body location independent activity monitoring," in *Proc. 9th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2016, pp. 190–197.

[45] F. Taia Alaoui, H. Fourati, A. Kibangou, B. Robu, and N. Vuillerme, "Urban transportation mode detection from inertial and barometric data in pedestrian mobility," *IEEE Sensors J.*, early access, Mar. 12, 2021, doi: 10.1109/JSEN.2021.3065848.

[46] A. Ejupi, C. Galang, O. Aziz, E. J. Park, and S. Robinovitch, "Accuracy of a wavelet-based fall detection approach using an accelerometer and a barometric pressure sensor," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2150–2153.

[47] K. Leuenberger, R. Gonzenbach, E. Wiedmer, A. Luft, and R. Gassert, "Classification of stair ascent and descent in stroke patients," in *Proc. 11th Int. Conf. Wearable Implant. Body Sensor Netw. Workshops*, Jun. 2014, pp. 11–16.

[48] K. Sankaran, M. Zhu, X. F. Guo, A. L. Ananda, M. C. Chan, and L.-S. Peh, "Using mobile phone barometer for low-power transportation context detection," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2014, pp. 191–205.

[49] M. Monteiro and A. C. Marti, "Using smartphone pressure sensors to measure vertical velocities of elevators, stairways, and drones," 2016, *arXiv:1607.00363*.

[50] M. Wu, P. H. Pathak, and P. Mohapatra, "Monitoring building door events using barometer sensor in smartphones," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2015, pp. 319–323.

[51] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016786551830045X

[52] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018.

[53] F. Attal *et al.*, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.

[54] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13029–13040, Jun. 2021.

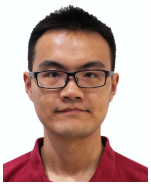
[55] SUTD. (2016). *National Science Experiment*. Singapore. Accessed: Feb. 22, 2022. [Online]. Available: https://www.family.sg/singapore-national-science-experiment.html
[56] E. Wilhelm *et al.*, "Wearable environmental sensors and infrastructure for mobile large-scale urban deployment," *IEEE Sensors J.*, vol. 16, no. 22, pp. 8111–8123, Nov. 2016.
[57] Y. Zhou, J. Wang, P. Shi, D. Dahlmeier, N. Tippenhauer, and E. Wilhelm, "Power-saving transportation mode identification for large-scale applications," 2017, *arXiv:1701.05768*.
[58] Skyhook. (2021). *Skyhook Location Services*. Accessed: Feb. 22, 2022. [Online]. Available: https://www.skyhook.com/
[59] J. He, Q. Zhang, L. Wang, and L. Pei, "Weakly supervised human activity recognition from wearable sensors by recurrent attention learning," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2287–2297, Mar. 2019.
[60] S. Hyuga, M. Ito, M. Iwai, and K. Sezaki, "Estimate a user's location using smartphone's barometer on a subway," in *Proc. 5th Int. Workshop Mobile Entity Localization Tracking GPS-Less Environments*, Nov. 2015, p. 2.
[61] H. Xia, X. Wang, Y. Qiao, J. Jian, and Y. Chang, "Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning," *Sensors*, vol. 15, no. 4, pp. 7857–7877, 2015.
[62] H. Ye, T. Gu, X. Tao, and J. Lu, "Scalable floor localization using barometer on smartphone," *Wireless Commun. Mobile Comput.*, vol. 16, no. 16, pp. 2557–2571, Nov. 2016.
[63] B.-J. Ho, P. Martin, P. Swaminathan, and M. Srivastava, "From pressure to path: Barometer-based vehicle tracking," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environments*, 2015, pp. 65–74.
[64] M. B. Del Rosario *et al.*, "A comparison of activity classification in younger and older cohorts using a smartphone," *Physiol. Meas.*, vol. 35, no. 11, p. 2269, 2014.
[65] M. Tachikawa, T. Maekawa, and Y. Matsushita, "Predicting location semantics combining active and passive sensing with environment-independent classifier," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2016, pp. 220–231.
[66] F. Haque, V. Dehghanian, A. O. Fapojuwo, and J. Nielsen, "A sensor fusion-based framework for floor localization," *IEEE Sensors J.*, vol. 19, no. 2, pp. 623–631, Jan. 2019.
[67] F. Zhao, H. Luo, X. Zhao, Z. Pang, and H. Park, "HYFI: Hybrid floor identification based on wireless fingerprinting and barometric pressure," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 330–341, Feb. 2017.
[68] Z. Wang, M. Jiang, Y. Hu, and H. Li, "An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 691–699, Jul. 2012.
[69] Y. Nam and J. W. Park, "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 2, pp. 420–426, Mar. 2013.
[70] C. Wang *et al.*, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2302–2311, Dec. 2016.
[71] P. Anastasopoulou, M. Tansella, J. Stumpp, L. Shammas, and S. Hey, "Classification of human physical activity and energy expenditure estimation by accelerometry and barometry," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 6451–6454.
[72] Z. Xu, J. Wei, J. Zhu, and W. Yang, "A robust floor localization method using inertial and barometer measurements," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2017, pp. 1–8.
[73] N. Kronenwett, S. Qian, K. Mueller, and G. F. Trommer, "Elevator and escalator classification for precise indoor localization," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2018, pp. 1–8.
[74] A. Dimri, H. Singh, N. Aggarwal, B. Raman, D. Bansal, and K. K. Ramakrishnan, "RoadSphygmo: Using barometer for traffic congestion detection," in *Proc. 8th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2016, pp. 1–8.
[75] C. Lang and S. Kaiser, "Classifying elevators and escalators in 3D pedestrian indoor navigation using foot-mounted sensors," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2018, pp. 1–7.
[76] R. Ichikari, L. C. M. Ruiz, M. Kourogi, T. Kurata, T. Kitagawa, and S. Yoshii, "Indoor floor-level detection by collectively decomposing factors of atmospheric pressure," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2015, pp. 1–11.
[77] Y. Li, Z. Gao, Z. He, P. Zhang, R. Chen, and N. El-Sheimy, "Multi-sensor multi-floor 3D localization with robust floor detection," *IEEE Access*, vol. 6, pp. 76689–76699, 2018.
[78] G. Zheng, "A novel attention-based convolution neural network for human activity recognition," *IEEE Sensors J.*, vol. 21, no. 23, pp. 27015–27025, Dec. 2021.
[79] K. Hamrick and K. J. Shelley, "How much time do Americans spend preparing and eating food?" in *Amber Waves:The Economics of Food, Farming, Natural Resources, and Rural America*. Washington, DC, USA: US Department of Agriculture (USDA), Economic Research Service, 2005. [Online]. Available: https://EconPapers.repec.org/RePEc:ags:uersaw:127415
[80] B. Monnot, F. Benita, and G. Piliouras, "How bad is selfish routing in practice?" 2017, *arXiv:1703.01599*.
[81] A.-L. Barabási, *Bursts: Hidden Patterns Behind Everything We Do, From Your e-Mail to Bloody Crusades*. Baltimore, MD, USA: Penguin, 2010.
[82] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
[83] E. J. Willemse, B. Tunçer, and R. Bouffanais, "Identifying highly dense areas from raw location data," in *Proc. 24th Int. Conf. Assoc. Comput.-Aided Architectural Design Res. Asia*, vol. 2, 2019, pp. 805–814.
[84] XGBoost. (2020). *XGBoost Documentation*. Accessed: Feb. 22, 2022. [Online]. Available: https://xgboost.readthedocs.io/
[85] A. Jain. (2020). *XGBoost Parameter Tuning Guide to Parameter Tuning in XGBoost*. Accessed: Feb. 22, 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/
[86] T. Yiu. (2019). *Understanding Random Forest*. Accessed: Feb. 22, 2022. [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2
[87] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.
[88] A. Sharma. (2021). *Gaussian Naive Bayes With Hyperparameter Tuning*. Accessed: Feb. 22, 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/01/gaussian-naive-bayes-with-hyperpameter-tuning/
[89] J. Friedman *et al.*, *The Elements of Statistical Learning* (Springer Series in Statistics) vol. 1, no. 10. New York, NY, USA: Springer, 2001.
[90] D. F. Young, B. R. Munson, T. H. Okiishi, and W. W. Huebsch, *A Brief Introduction to Fluid Mechanics*. Hoboken, NJ, USA: Wiley, 2010.

Ajaykumar Manivannan received the B.Eng. degree in aerospace engineering from Hindustan University, India, in 2013, and the M.Sc. degree in aerospace engineering from Technische Universität München (TUM), Germany, and Nanyang Technological University (NTU), Singapore, in 2016. He is currently pursuing the Ph.D. degree with the Department of Mechanical Engineering, University of Ottawa, Ottawa, ON, Canada. His research interests include computational social science, human mobility, network science, and fluid mechanics.

Elias J. Willemse received the B.Eng. and Ph.D. degrees in industrial engineering from the University of Pretoria, Pretoria, South Africa, in 2007 and 2016, respectively. He was a Senior Lecturer at the University of Pretoria, Pretoria, from 2016 to 2019. He was a Postdoctoral Researcher with the Singapore University of Technology and Design, Singapore, from 2018 to 2020. He is currently the CTO of Waste Labs. His research interests include optimization modeling, and their application to transportation and waste collection systems.

Balamurali B. T. received the Ph.D. degree in electrical and computer engineering from the University of Auckland, New Zealand, in 2015. After his Ph.D. degree, he worked as a Researcher at the Gastro Intestinal Group, Auckland Bioengineering Institute. He was also a Lecturer at the Auckland University of Technology. Prior to his Ph.D. endeavor, he was a Senior Design and a Development Engineer at Tata Elxsi, India. He is currently a Postdoctoral Research Fellow working at the Singapore University of Technology and Design. He is passionate about artificial intelligence and trying to solve a variety of problems related to bioacoustics processing, detection and classification of bio-signals, automatic speech/speaker recognition, spoofed-speech detection, blacklisted speaker identification, blind source separation, music classification, fluid flow classification, and early-dementia prediction.

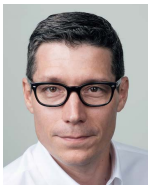
Wei Chien Benny Chin received the B.Sc., M.Sc., and Ph.D. degrees in geography from the National Taiwan University, Taiwan, in 2011, 2014, and 2018, respectively. He is currently a Postdoctoral Researcher with the Singapore University of Technology and Design. He is also participating in projects associated with complex human movement and the horizontal and vertical dimensions of spatial structures. His research interests include space-time analysis and complex network analysis.



Yuren Zhou received the B.Eng. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2014, and the Ph.D. degree in engineering from the Singapore University of Technology and Design, Singapore, in 2019. His research interests include data mining, machine learning, and their applications in human mobility, building energy management, and the Internet of Things.

Bige Tunçer received the B.Arch. degree from Middle East Technical University in 1993, the M.Sc. degree from Carnegie Mellon University in 1996, and the Ph.D. degree from the Delft University of Technology in 2009. She is now an Associate Professor at the Singapore University of Technology and Design, where she leads the Informed Design Laboratory. The lab's research focuses on data driven architectural and urban design. She leads and participates in large multi-disciplinary research projects in evidence informed design, the IoT, and big data.

Alain Barrat received the B.Sc. and M.Sc. degrees in physics from École Normale Supérieure, Paris, France, in 1992 and 1994, respectively, and the Ph.D. degree in theoretical physics from UPMC University, Paris, in 1996. He was a Postdoctoral Fellow at the Abdus Salam ICTP, Trieste, Italy, and entered the French National Council for Scientific Research (CNRS) as a Junior Researcher in 1998. He is currently a CNRS Senior Researcher at the Centre de Physique Théorique, Marseille, France. Since April 2019, he has been a Specially Appointed Professor at the Tokyo Tech World Research Hub Initiative, Tokyo, Japan. His research interests include statistical physics and its interdisciplinary applications. He is an expert in the field of complex networks, from fundamental aspects to applications ranging from computational social science to epidemiology.



Roland Bouffanais (Member, IEEE) received the B.Sc. and M.Sc. degrees in physics from École Normale Supérieure (ENS Lyon), Lyon, France, in 1997 and 1999, respectively, the M.Sc. degree in physics from UPMC Paris Sorbonne University, Paris, France, in 1999, and the Ph.D. degree in engineering from the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2007. He was a Postdoctoral Fellow and an Associate with the Department of Mechanical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He is now an Associate Professor with the Department of Mechanical Engineering, University of Ottawa, Ottawa, ON, Canada. His interdisciplinary research spans a vast breadth of areas and focused on the design and control of decentralized complex systems, multi-agent systems, leader–follower consensus dynamics, and non-linear dynamical systems.

Part III

What are the patterns of vertical mobility in a city?

6

Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility

This chapter includes a manuscript under preparation for submission to Nature Cities, which addresses the third research question (R3) posed in Chapter 1, focusing on unraveling the patterns of vertical mobility and its interplay with horizontal mobility based on the vertical mobility data set derived using the methodology developed in Chapter 5.

Manuscript details :

- Manivannan, A., Willemse, E. J., Chin, B., Bige, T., Barrat, A., & Bouffanais, R. Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility (under preparation for Nature Cities)

Summary: Vertical mobility—movement in the built environment's vertical dimension via elevators, escalators, and stairs—is a vital yet often overlooked component of urban mobility. Traditional studies focus predominantly on horizontal movement, almost always neglecting the significance of vertical mobility in high-density, multi-level cities. This study introduces vertical mobility into urban mobility analysis, utilizing real-world data from over 20,000 individuals in Singapore. We quantify the prevalence of vertical mobility, characterize its complexity through heavy-tailed distributions of key mobility metrics, and examine its interplay with horizontal mobility. Our findings reveal that vertical mobility occurs as frequently as horizontal movement and exhibits significant variability and heterogeneity, thereby indicating complex underlying patterns. The moderate but significant correlation between vertical and horizontal mobility underscores their interdependence. These results highlight the necessity of incorporating vertical mobility into urban planning and transportation models to create sustainable, efficient, and resilient cities that fully embrace their three-dimensional nature.

Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility

Ajaykumar Manivannan[1], Elias J. Willemse[2], Wei Chien Benny Chin[3], Bige Tunçer[4], Alain Barrat[5], Roland Bouffanais[1,6*]

[1*]Department of Mechanical Engineering, University of Ottawa, Ottawa, K1N6N5, Ontario, Canada.
[2]Waste Labs Pte. Ltd., 059915,Singapore.
[3]Department of Geography, National University of Singapore, 117568, Singapore.
[4]Information Systems in the Built Environment, The Eindhoven University of Technology, City, 5612 AZ, North Brabant, Netherlands.
[5]Center for Theoretical Physics, CNRS and Aix Marseille University, Marseille, 13009, Provence-Alpes-Côte d'Azur, France.
[6]Department of Computer Science & Global Studies Institute, University of Geneva, Genève CH-1211, Switzerland.

*Corresponding author(s). E-mail(s): roland.bouffanais@unige.ch;

Abstract

Vertical mobility—movement in the built environment's vertical dimension via elevators, escalators, and stairs—is a vital yet often overlooked component of urban mobility. Traditional studies focus predominantly on horizontal movement, neglecting the significance of vertical mobility in high-density, multi-level cities. This study introduces vertical mobility into urban mobility analysis, utilizing real-world data from over 20,000 individuals in Singapore. We quantify the prevalence of vertical mobility, characterize its complexity through heavy-tailed distributions of key mobility metrics, and examine its interplay with horizontal mobility. Our findings reveal that vertical mobility occurs as frequently as horizontal movement and exhibits significant variability and heterogeneity, indicating complex underlying patterns. The moderate but significant correlation between vertical and horizontal mobility underscores their interdependence. These results highlight the necessity of incorporating vertical mobility into urban planning and

transportation models to create sustainable, efficient, and resilient cities that fully embrace their three-dimensional nature.

Keywords: Human Mobility, Volumetric Urbanism, Urban Mobility, Multi-layer Transportation, Vertical Movement

1 Introduction

Human mobility in cities is a cornerstone of sustainable urban development, affecting nearly every aspect of city life from economic activity and social equity to environmental sustainability and public health. Understanding mobility patterns is essential for urban planning, transportation design, and policy-making. Traditionally, large-scale mobility analyses have been limited to two-dimensional urban settings, focusing solely on horizontal movement across the urban landscape [1–3]. This approach overlooks the vertical dimension of urban spaces, which is increasingly significant in high-density cities characterized by vertical growth and complex multi-level infrastructures.

Vertical mobility—movement in the built environment's vertical dimension via elevators, escalators, and stairs—is integral to intra-urban human mobility. In cities like Singapore, vertical transportation is deeply embedded in daily life. The Singapore government estimates that the city has over 70,000 passenger lifts and 7,000 escalators installed [4]. Moreover, 95% of Singapore's population lives in low- to high-rise apartments, with only 5% residing in landed properties [5]. Most train stations are elevated or underground, and all are equipped with escalators and elevators, often integrated into multi-floor shopping centers. Despite this, the vertical component of mobility remains underrepresented in urban mobility studies.

The consequences of overlooking the vertical dimension in urban mobility are significant. A two-dimensional analysis may effectively integrate vertical displacements into horizontal ones, potentially obscuring bottlenecks at vertical-horizontal junctions or neglecting the need for specific infrastructure to facilitate vertical movement. When considering multiple options to travel from one point to another, a two-dimensional perspective is insufficient to capture the complexity of possible paths, especially in densely built environments where vertical connectivity plays a crucial role.

Accounting for the influence of the third dimension is paramount to shedding light on the intricate interplay between vertical and horizontal mobilities, including their effects on the transport layer. Such a comprehensive mobility analysis can profoundly impact the design of sustainable cities, inform infrastructure development, and enhance the efficiency of transportation systems.

Despite the prevalence of vertical mobility in daily life, we lack quantitative knowledge about how much verticality individuals experience daily. Previous studies have provided qualitative or theoretical insights [6], and geographers and city councils have mapped building heights accessible through commercial maps [7, 8]. Some urban planners and designers have begun to study urban layouts as three-dimensional networks [9, 10]. However, there is a gap in the empirical quantification of vertical mobility experiences based on real-world data.

Furthermore, the heavy-tailed distributions of human mobility metrics in horizontal movement reveal underlying heterogeneity in factors influencing human mobility [11, 12]. These distributions are fundamental mechanisms driving human mobility and have implications for urban processes such as traffic congestion, resource allocation, and the spread of infectious diseases [13, 14]. Whether vertical mobility patterns exhibit similar complexity and heterogeneity remains an open question.

This study aims to address the following research questions:

1. How prevalent is vertical mobility in urban environments, and what patterns characterize its magnitude and frequency?
2. Do vertical mobility patterns exhibit complexity and heterogeneity similar to horizontal mobility, as evidenced by heavy-tailed distributions of key mobility metrics?
3. What is the interplay between vertical and horizontal mobility, and how are they interconnected in daily urban movement?

To answer these questions, we conduct a comprehensive analysis of vertical mobility using real-world data collected from a large population in Singapore. We quantify the extent of vertical mobility experienced by individuals, analyze the distributions of displacement, waiting time, and mobility duration, and examine the interplay between vertical and horizontal movements. By integrating vertical mobility into the study of urban movement patterns, we aim to provide empirical evidence of its significance and advocate for its inclusion in urban mobility models and planning.

Our findings reveal that vertical mobility is a prevalent and integral component of daily urban life, occurring with comparable frequency to horizontal movements. We observe that vertical mobility patterns exhibit heavy-tailed distributions, indicating complexity and heterogeneity similar to horizontal mobility. Moreover, there is a significant interplay between vertical and horizontal mobility, underscoring the need to consider them jointly in urban mobility analyses. These results have profound implications for urban planning, transportation design, and sustainability, highlighting the importance of adopting a three-dimensional perspective in understanding and modeling human mobility in cities.

2 Results

Prevalence and Patterns of Vertical Mobility

We examine the vertical dimension of city living and quantify how much verticality residents of Singapore experience in their homes, schools, and public spaces (referred to hereafter as a free-range environment). In addition, to gain a comprehensive understanding of human mobility patterns in urban environments, we analyzed the prevalence and patterns of vertical mobility and compared them with horizontal movement. This comparison highlights the significance of vertical displacement in daily activities and its role alongside horizontal mobility.

Vertical Urban Living: Heights experienced through vertical movement The height of the built environment is inferred from the maximum vertical displacement of a subject in a given built environment type (home, school, or free-range). Singapore's city living is highly vertical for nearly all residents. The median home height is around 17 meters (approximately six floors), indicating elevated living (Fig. 1B). The top 25^{th} percentile lives above ten floors. Residential floor height is consistent across all age groups (Ages 6–21), meaning families with young children (Ages 6–11) are just as likely to live at any floor level as families with older children (Ages 17+) (Refer to Supplementary Fig. S3). In free-range environments commonly encountered by all Singaporeans, the median height is about 9 meters (2 to 3 floors), showing that residents regularly experience elevated spaces, like train platforms above or below ground. Among the studied population, 95% live at least one floor above the surface (90% if considering all eligible subjects), and 80% navigate elevated urban environments to move through the city.

Magnitude: The total vertical displacement is significantly less than horizontal movement in terms of distance (refer to Sec. 3 for discussion on vertical mobility's non-trivial energy costs). A median individual travels approximately 78 ± 32 meters (median $\pm$ median absolute deviation) vertically per day (102 ± 33 meters for the adult population) and spends about 7 ± 4 minutes doing so. In contrast, in the horizontal dimension, a median individual travels approximately 11 ± 8 kilometers daily (26 ± 12 kilometers for the adult population) and spends about 1.5 ± 0.75 hours. Although vertical displacement covers shorter distances, it represents a crucial component of daily mobility.

Frequency: Vertical and horizontal mobility occur with comparable frequency. A median individual engages in vertical movement approximately 7 ± 2 times per day, equating to 0.78 ± 0.22 times per hour. Horizontally, a median individual moves about 6 times per day, or 0.63 ± 0.17 times per hour. This indicates that vertical mobility is a high-frequency event, integral to daily routines. The timeline of vertical mobility events (Fig. 2D) underscores this pattern, showing continuous vertical movement throughout the day.

Temporal Patterns: Vertical and horizontal mobility exhibit similar temporal patterns (Fig. 2D). Three peaks are observed at approximately 7 a.m., 10 a.m., and 1 p.m., corresponding to school start time, break periods, and school dismissal, respectively. Additionally, increased vertical displacement occurs just before school begins

(around 6 a.m.) and after school ends (1–4 p.m.), associated with movements in residential and public environments. In the general population, these peaks likely shift to 7–8 a.m. for the morning and 5–6 p.m. for the evening. The synchronization of these patterns indicates that vertical and horizontal mobility are intertwined aspects of daily activity cycles, a relationship explored further in subsequent results.

Location-Specific Patterns: The majority of total vertical displacement occurs at home (42%), even though it accounts for only 24% of all vertical movements. Conversely, schools contribute to 31.5% of the total vertical displacement but are the sites of the most vertical movements (43.4%). Notably, approximately one-third (32.6%) of all vertical movements take place in public or free-range areas, highlighting the level of verticality potentially experienced by the general population in their everyday surroundings. This distribution emphasizes the importance of various environments in shaping vertical mobility patterns.

In summary, while individuals travel shorter distances vertically than horizontally each day, they engage in vertical movements just as frequently and at similar times. Vertical mobility is thus a constant and significant part of daily routines, underscoring its role in urban mobility dynamics.

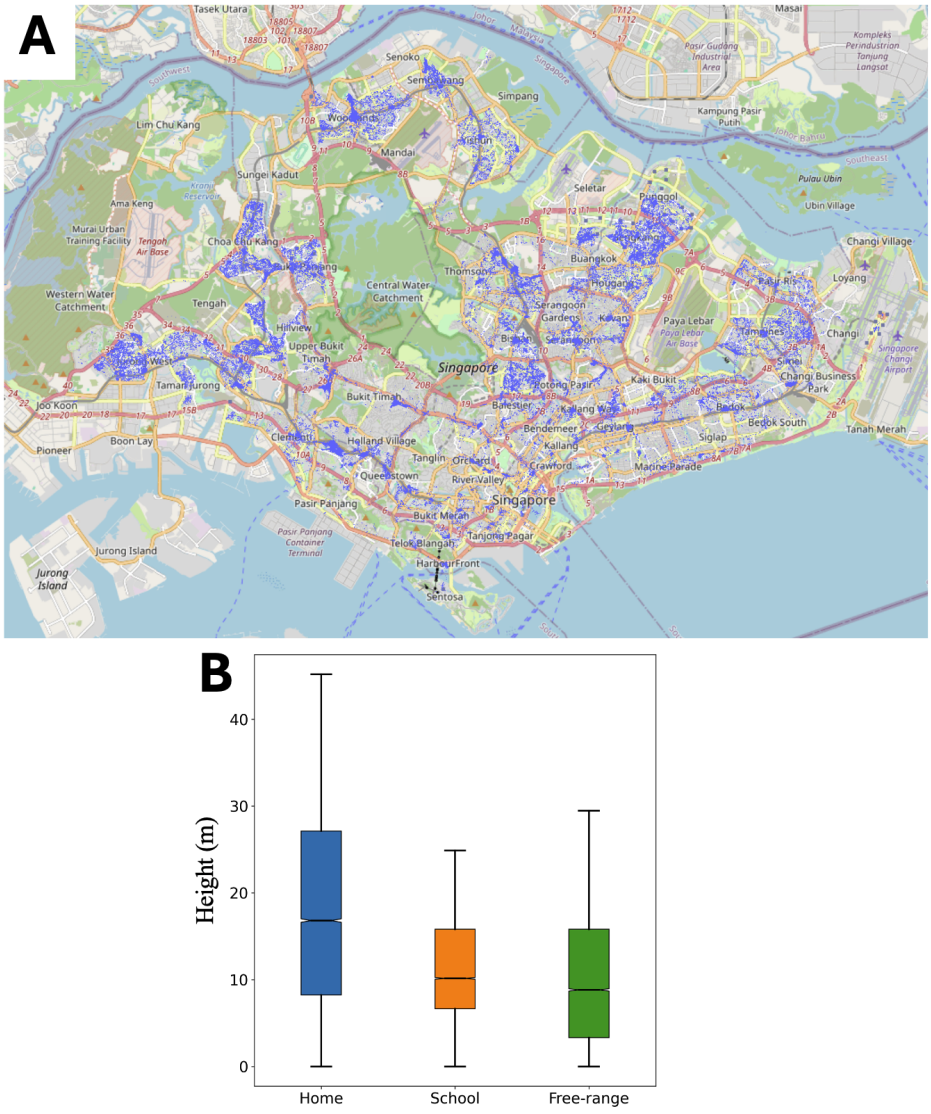


Fig. 1 Spatial patterns of vertical human mobility reveal that city living is highly vertical. A, Location (blue dots) of all vertical mobility events in Singapore recorded by NSE 2016 dataset. The spatial map shows comprehensive coverage of Singapore's built environment based on $\approx 390,000$ upward or downward only vertical movements of $\approx 20,000$ subjects. **B,** Boxplot of the inferred height of different built environment types (home, school, and free-range) reveals that residents experience significant verticality in daily life. The whiskers represent the 5^{th} and 95^{th} percentile of the data, and the notch represents the median.

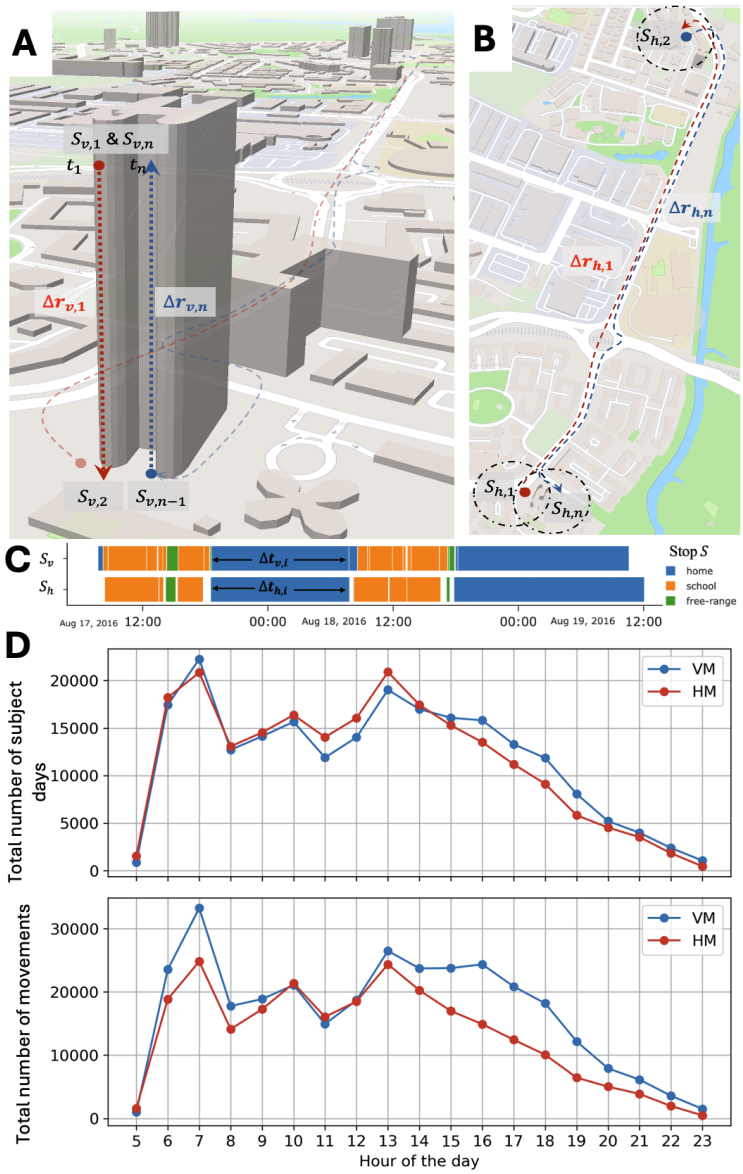


Fig. 2 Mobility illustration and temporal patterns A, Illustration of vertical displacement Δr_v and stop S_v along the vertical plane. **B,** Illustration of horizontal displacement Δr_h and stop S_h along the horizontal plane. **C** Timeline of movement and stops for a sample subject. White block represents movement. Vertical and horizontal movement is observed to take place very close to each other. **D,** Timeline of the total number of movements summed for the entire population by hour of the day. Horizontal and vertical movement exhibit similar temporal patterns.

Heavy-Tailed Distributions Reveal Complexity in Vertical Mobility patterns

Understanding the complexity inherent in vertical mobility patterns is essential for urban planning, public transportation, and social dynamics in densely populated cities. We interpret heterogeneity—variability within the distribution of a given mobility metric—as an indication of complexity in observed mobility patterns. Specifically, when a distribution is heavy-tailed and spans several orders of magnitude (or reaches the maximum possible range), it suggests the presence of a complex underlying process that warrants further investigation. In this study, we examine three fundamental mobility metrics central to human mobility [13]: **(a) Displacement** Δr, defined as the cumulative distance of a trip; **(b) Waiting time** Δt, defined as the time between trips or time spent at the destination of a trip; and **(c) Mobility duration** τ, defined as the time elapsed during a trip. For a comprehensive list of human mobility metrics, we refer the reader to a detailed review by Barbosa et al. [13]. To evaluate the heavy-tailedness and heterogeneity of these distributions, we consider five candidate models, ranging from the power law (most variability) to the exponential distribution (least variability), as detailed in the Methods section.

Displacement Δr: The distribution of vertical displacement $P(\Delta r_v)$ is best fitted by a stretched exponential function of the form $f(x) = (\Lambda x)^{\beta-1} e^{-(\Lambda x)^{\beta}}$ with a stretching exponent $\beta = 0.72 \pm 0.06$ (denoting mean $\pm$ standard deviation obtained from randomly drawing 100 subject's data 1000 times) (Fig. 3A). This value of β indicates that the distribution deviates from a pure exponential decay ($\beta = 1$), suggesting moderate heterogeneity in vertical displacement ($\beta = 0$ is pure power law). The distribution drops significantly after $\approx$ 50 meters (approximately 15 floors), reflecting physical constraints in building heights within the urban environment. In contrast, the distribution of horizontal displacement $P(\Delta r_h)$ is best described by a truncated power law of the form $f(x) = x^{-\alpha} e^{-\lambda x}$ with an exponent $\alpha = 1.15 \pm 0.03$ and a cutoff value (till which the fit is power law and after that turns exponential) $1/\lambda \approx 31 \pm 3$ km (Fig. 3B). The cutoff corresponds to the physical boundaries of Singapore, where the maximal distance between publicly accessible ends of the island city-state is approximately 40 km. The heavy-tailed nature of $P(\Delta r_h)$ spans four orders of magnitude, indicating significant variability and complexity in horizontal movements. Although $P(\Delta r_v)$ spans only two orders of magnitude, it still exhibits heterogeneity within the vertical limits of the urban landscape. This suggests that vertical displacement, while constrained by building heights, contributes to the overall complexity of mobility patterns.

Waiting Time Δt: The distribution of waiting time in the vertical plane $P(\Delta t_v)$ is well approximated by a stretched exponential with $\beta = 0.48 \pm 0.019$ (Fig. 3C), indicating a higher degree of heterogeneity than an exponential distribution. For the horizontal plane, $P(\Delta t_h)$ is best fitted by a truncated power law with $\alpha = 1.01 \pm 0.02$ and a cutoff at approximately 9 ± 0.5 hours (Fig. 3D). The stretched exponential with $\beta = 0.48$ is also a close fit for $P(\Delta t_h)$, suggesting similar underlying dynamics. Both $P(\Delta t_v)$ and $P(\Delta t_h)$ span comparable ranges and exhibit heavy-tailed behavior across three orders of magnitude. This indicates significant variability in waiting times, reflecting the complex temporal patterns of human mobility. Notably, the distributions

reveal natural temporal scales in mobility. There is a noticeable drop around 15 hours, corresponding to the maximum time between returning home and returning to school among the population. Additionally, slight peaks around 1 and 2 hours in $P(\Delta t_v)$ likely correspond to school break times during which individuals move vertically. Similar patterns are observed in $P(\Delta t_h)$, albeit more pronounced. These findings suggest that vertical and horizontal waiting times share similar distribution patterns, reinforcing the notion that mobility in both planes is intertwined.

Mobility Duration τ: The distribution of vertical mobility time $P(\tau_v)$ is best fitted by a truncated power law with an exponent $\alpha = 1.68 \pm 0.05$ and a cutoff at $1/\lambda \approx 13$ minutes (Fig. 3E). This indicates moderate heterogeneity and reflects the limited duration of vertical movements, such as ascending or descending within buildings. Similarly, the distribution of horizontal mobility time $P(\tau_h)$ is best described by a truncated power law with $\alpha = 1.00 \pm 0.0007$ and a cutoff at $1/\lambda \approx 2$ hours (Fig. 3F). This suggests a higher degree of variability in horizontal movement durations, possibly due to varying travel distances and modes of transportation.

The heavy-tailed distributions of $P(\Delta r_v)$, $P(\Delta t_v)$, and $P(\tau_v)$ indicate that vertical mobility exhibits complex properties characterized by significant variability within natural boundaries and time scales. Notably, the similarities in the distributions and exponents of $P(\Delta t_v)$ and $P(\Delta t_h)$ suggest that vertical and horizontal mobility patterns are potentially interconnected. This interplay warrants further investigation, which we explore in detail in the next section.

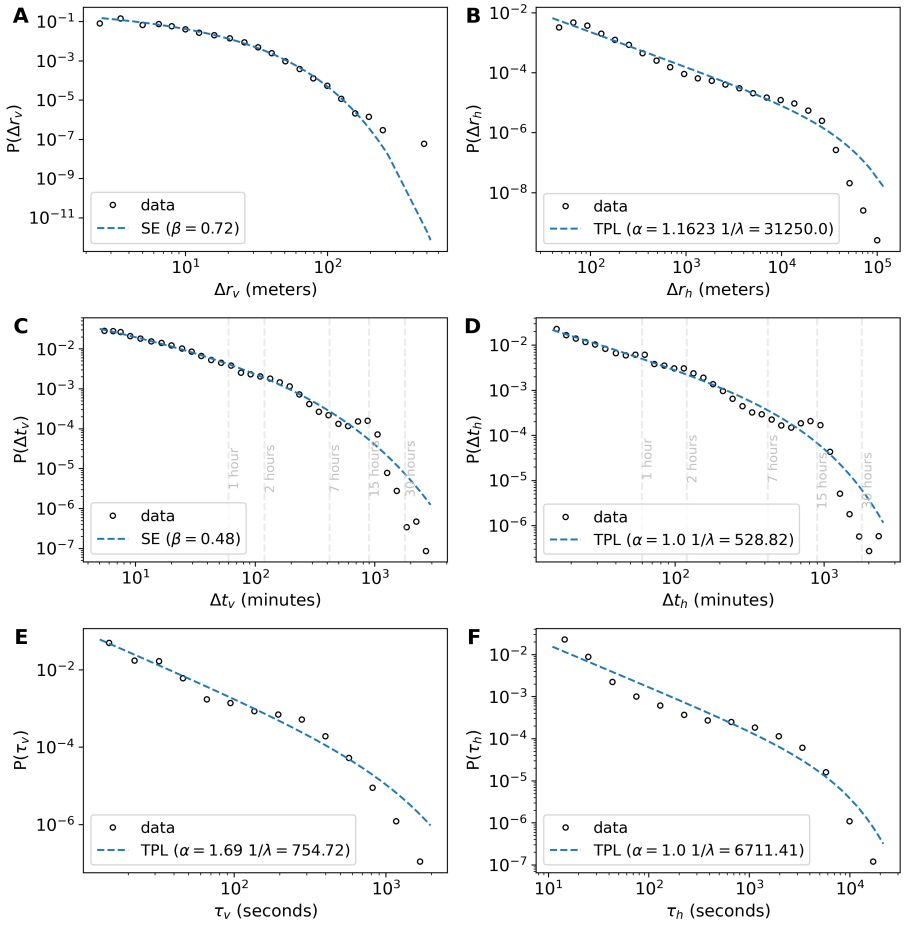


Fig. 3 Vertical human mobility patterns show heterogeneity. SE: Stretched Exponential; TPL: Truncated Power Law. All graphs are in a log-log plot with logarithmic binning. The black circles denote the mobility properties' probability density function $P(x)$ for $x > x_{min}$ (for all x_{min}, refer to Supplementary Table S2). $P(x)$ for **A & B,** displacement $\triangle r$, **C & D,** waiting time $\triangle t$, and **E & F,** mobility duration τ for respective mobility dimensions.

Interplay of Vertical and Horizontal Mobility

Having analyzed the magnitude and frequency of vertical movement, we now explore the role of the vertical transportation layer in urban mobility within Singapore. Specifically, we investigate the interplay between vertical and horizontal mobility to understand how they are interconnected.

Transition Count: A simple measure of this interplay is the number of transitions between horizontal and vertical mobility (regardless of waiting time between movements). The results show that approximately 75% of all transitions occur between horizontal and vertical mobility, while only 25% occur between movements of the same type (Table 1). This is unsurprising in cities like Singapore, where transitioning between the two mobility dimensions is seamless and continual.

Correlation Between Daily Vertical and Horizontal Movement for the Population: The daily total vertical displacement and the number of vertical movements for the population studied are moderately correlated with the total horizontal displacement and the number of horizontal movements. To visualize this relationship, we show the binned scatter plots where the conditional mean and standard deviation of vertical mobility properties are plotted against the equal-sized (approximately) bins of horizontal mobility.

The results indicate that the mean total vertical displacement increases gradually with increasing ranges of total horizontal displacement and number of horizontal movements. The Spearman's correlation coefficient r_s between the total vertical displacement and total horizontal displacement is approximately 41%*** (Fig. 4A), and 20%*** between total vertical displacement and number of horizontal movements (Fig. 4B). Similarly, the number of vertical movements increases with increasing bins of total horizontal displacement (Fig. 4C) and number of horizontal movements (Fig. 4D, with $r_s \approx$ 47%*** and 32%***, respectively. These moderate correlations suggest a significant interplay between the two types of mobility.

Furthermore, when considering only movements in public spaces, we observe similar or slightly higher correlation values of 45%*** and 51%*** between the respective sums and counts of mobility dimensions (Fig. 4E&H). In the school environment, where only shorter horizontal travels are possible (maximum diameter of schools ranging from approximately 160 to 500 meters depending on age group), the correlation values remain between $r_s \approx$ 34%*** and 50%*** (Fig. 4I–L). The correlation coefficients are mostly consistent when using both Pearson and Spearman methods, with some exceptions in the school environment (between total horizontal displacement and sum & count of vertical displacement). Importantly, our results demonstrate that more frequent and longer horizontal mobility moderately co-occurs with more frequent and longer vertical mobility. This consistency holds even when the spatial scale of mobility is short (schools) or typically long (public spaces). However, as seen by the high standard deviation (shaded areas of Fig. 4) and the moderate correlation values, vertical mobility activities are also moderately independent of horizontal mobility activities. Our results contrast with Rode et al. [6], who compared the simulated horizontal and vertical mobility data and stated that no clear relationship exists between them.

These findings show that vertical mobility is highly intertwined with horizontal mobility, underscoring its role as a critical dimension in urban movement. This emphasizes the necessity of exploring vertical mobility as a vital transportation layer in multimodal urban mobility systems.

Table 1 Number of transitions between vertical (VM) and horizontal mobility (HM). 75% of transitions happen between mobility in the horizontal and vertical planes.

Type	Number	Ratio
VM to HM	255,718	37.8 %
HM to VM	249,308	37.9 %
VM to VM	57,122	8.4 %
HM to HM	113,354	16.7 %

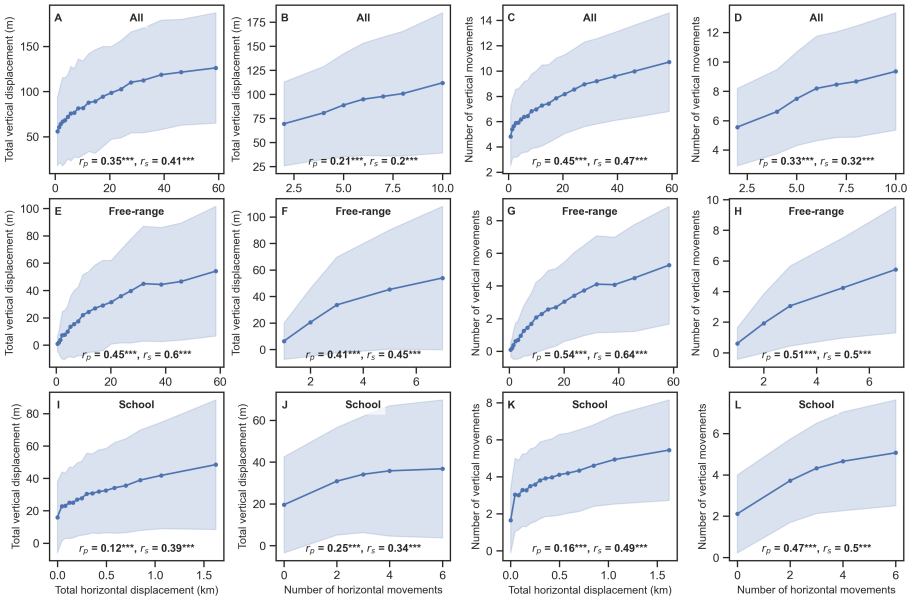


Fig. 4 Daily movement statistics of subjects reveal horizontal and vertical mobility to be intertwined. Statistical significance of correlation: $***$: $p < 0.001$, $**$: $p < 0.01$, $*$: $p < 0.05$). The binned scatter plots show the conditional mean (blue dots) and standard deviation (shaded region) of vertical mobility properties (y-axis) plotted against the equal-sized (approximately) bins of horizontal mobility properties (x-axis). The mobility properties shown are the daily total displacement and number of mobility events for vertical and horizontal. **A**–**D,** All environments, **E**–**H,** Free-range environment, and **I**–**L,** School environment (movement within school).

3 Discussion

Vertical mobility's critical role is established through its prevalence, complexity, and integration. In this article, we explore its implications for cities regarding urban planning, urban commuting, modeling and prediction of mobility, energy and sustainability, and multi-layer transportation.

Volumetric Urbanism: High-density cities experiencing upward growth require a new paradigm in urban planning. While current cities benefit from planned and emergent street networks that are well-connected for seamless horizontal movement across urban spaces, future urban environments will necessitate enhanced connectivity across floor levels and between buildings on a larger scale than currently present—for example, in integrated developments [15]. This necessitates studying cities as volumetric spaces. Our study contributes to the concept of volumetric urbanism by providing empirical support for moving beyond horizontalism in urban mobility analysis [16–20]. Specifically, by quantifying the actual heights experienced by individuals through data-driven methods, we distinguish our work from simulations based solely on building heights that do not account for individual experiences [6].

Urban Commuting: Fundamental daily travel statistics are crucial for understanding and improving urban commuting. We present descriptive statistics on daily spatial and temporal movement patterns in both vertical and horizontal dimensions. Our horizontal mobility results align with those reported by government agencies [5] and existing literature for the general population in Singapore [21]. This similarity between the student population studied and the broader population supports the generalizability of our vertical mobility findings. This is further reinforced by the fact that both groups navigate the same built environment where significant vertical transportation is involved—namely, homes and free-range environments. Additionally, our vertical mobility daily travel statistics are comparable to the estimates reported by Rode et al. [6], where the average vertical distance traveled per person per day to the center of a sample area ($500\,\mathrm{m} \times 500\,\mathrm{m}$) in Singapore ranges from 90 to 150 meters, depending on the region (based on simulated data from known building heights). For Hong Kong, this value is reported as 150 meters, and for Kuwait, 30 meters [6].

Energy Costs, Sustainability, and Carbon Footprint: Frolking et al. [22] reported data-driven findings on the shift toward verticality in urban landscapes since the 1990s, which has significant consequences for energy costs and harmful emissions associated with developing vertically built environments. However, the energy costs related to urban movement due to verticality are often not attributed to human mobility but are instead reported as part of building energy consumption. The contribution of vertical transportation to a city's total energy consumption can be non-trivial, with reports indicating it accounts for 0.7% of total energy consumption in the European Union [23] and 1.5% in New York City attributed to elevators alone [24].

While the magnitude of vertical displacement is a fraction of its horizontal counterpart—specifically, the median individual moves horizontally 141 times farther than vertically—the difference in energy consumption between vertical and horizontal transportation is not proportionally as vast. Based on our findings on daily vertical movement and energy estimates from the literature, the daily energy consumed for horizontal movement is estimated to be roughly 18 to 37 times greater than that

for vertical movement (see Supplementary Information for details). When considering the displacement difference of 141 times, this indicates that vertical mobility, despite covering shorter distances, has a disproportionately higher energy cost per unit distance.

Moreover, the energy consumption due to vertical movement is likely to have a proportional effect on carbon emissions, thereby increasing the carbon footprint of human movement beyond current estimates. Consequently, vertical mobility, due to its non-trivial energy costs, is likely to have a considerable impact on urban sustainability and carbon footprints. Recognizing and addressing the energy demands of vertical transportation is essential for developing strategies aimed at reducing emissions and promoting sustainable urban development.

Human Mobility Models: The characterization of empirical mobility patterns is crucial for modeling, reproducing, and predicting human movement. In this study, we contribute to the state-of-the-art in this area by characterizing the distributions of displacement, waiting time, and mobility duration for vertical mobility, while benchmarking our findings against horizontal mobility patterns. Power-law functions are typically found to best fit the distribution $P(\Delta r_h)$ for horizontal displacement, although the reported exponent values vary widely ($\alpha = 1$ to 3) [25]. For example, power-law exponents reported in the literature include: $\alpha = 1$ to 1.4 (Mobile Flow Records) [26], 1.39 to 1.57 (Global Positioning System GPS data) [27], 1.62 (Twitter data) [28], 1.05 (travel diaries) [29], 1.2 (Taxi GPS) [30], 1.35 to 1.82 (GPS data) [31], 1.55 (Call Detail Records CDR) [32], and 1.75 (CDR) [33]. However, for intra-urban mobility, exponential-like functions or truncated power laws—which behave like exponential functions beyond a cutoff point—are often found to provide a better fit due to the limiting nature imposed by city boundaries. Our results are consistent with this, as we find the truncated power law to be a highly suitable candidate, with exponent values within the reported range.

Additionally, the distribution of vertical displacement aligns with theoretical arguments and empirical evidence reported by Yan et al. [29], where $P(\Delta r)$ for a single mode of transportation follows an exponential-like form, attributed to travel cost being proportional to displacement (cost $\propto r$).

The peak times of mobility patterns observed are consistent with those reported in the literature [25]. Interestingly, we observed the daily rhythm of mobility behavior from a single transportation mode—vertical mobility—which closely follows the horizontal mobility patterns undertaken through multiple modes (bus, train, car, and walking).

Intra-urban Multi-layer Transportation: Waiting times for vertical transportation can be significant. At peak times in office towers and transit stations with limited lift capacity, bottlenecks can occur, creating wait-times comparable to those experienced when waiting for bus services (approximately $10 \sim 15$ minutes) in cities like Singapore. Given the high interplay and correlation between daily horizontal and vertical movements, it is necessary to consider the vertical layer as part of multi-layer transportation research. Moreover, wayfinding algorithms—which are already challenging to design for multimodal transportation—will become even more complex without sufficient knowledge and accounting of the vertical transportation

layer [34]. This interdisciplinary research, influenced by methodological developments in complex systems studies, comes at a time when transportation research is receiving increased attention and contributions from complex systems researchers, in addition to traditional transportation engineers and planners [35].

Limitations and Scope: While our study provides valuable insights, several limitations should be acknowledged. The dataset primarily comprises a specific demographic group (students aged 6–21), which may not fully represent the broader population's mobility patterns. Additionally, the study focuses on Singapore, a highly urbanized and compact city-state, which may limit the generalizability of the findings to other urban contexts with different geographical and socio-economic characteristics. Nevertheless, several key insights could, to an extent, be extended to other dense cities.

In conclusion, our study demonstrates that vertical mobility is an integral part and a complex aspect of urban movement, occurring with comparable frequency to horizontal mobility and exhibiting significant variability as evidenced by heavy-tailed distributions. The interplay between vertical and horizontal mobility highlights the critical need for a three-dimensional perspective in urban mobility analysis, urban planning, and transportation design. Recognizing the significant energy costs and sustainability implications associated with vertical mobility is essential for developing strategies to reduce emissions and enhance efficiency. Our findings advocate for the inclusion of vertical mobility in future research and urban models, emphasizing the importance of comprehensive datasets that capture the full spectrum of human movement. Embracing vertical mobility in urban studies is crucial for creating sustainable, efficient, and resilient cities that meet the evolving needs of their inhabitants in increasingly vertical urban environments.

4 Methods

4.1 Data

4.1.1 National Science Experiment 2016 Mobility Dataset

The National Science Experiment (NSE) 2016 dataset consists of 50,000 students aged between 6 and 21 years from 89 schools. Each student carried a wearable device called SENSg, which consisted of built-in environmental sensors, motion sensors, and communication units. These devices recorded and transmitted data related to the Activities of Daily Living (ADL) of the students to a cloud server [36]. Data was recorded every 13 to 18 seconds during weekdays (Monday to Friday) over an eight-week period in 2016, with each subject participating for up to one week. The device entered sleep mode to reduce battery consumption when the processed inertial measurement unit (IMU) signal indicated no user movement. The SENSg devices were distributed to students on Mondays and collected back on Fridays. Therefore, full-time scale daily ADL data is only available on Tuesdays, Wednesdays, and Thursdays.

The NSE 2016 data was curated to remove devices with missing location data exceeding 30% in a day or with less than 4 hours of time coverage. The description

of the curated mobility dataset is shown in Table 2. The curated NSE 2016 mobility dataset contains 20,570 subjects with 318,434 vertical displacements and 249,512 horizontal displacements measured over a maximum period of three days per subject. Of the 20,570 subjects, approximately 7,900 had mobility events for three days, and approximately 6,700 had data for two days. The dataset has relatively low time coverage because data was recorded only when the IMU sensor detected body movement. Nevertheless, the dataset has a median time coverage of 7.2 hours per day (refer to Supplementary Fig. S1 for a boxplot of time coverage per subject per day).

Table 2 NSE 2016 Mobility Dataset A: Maximum period of 3 days per subject.

Population	No. schools	No. subjects	No. VD	No. HD
All	85	20,570	318,434	249,512
Primary (6–11 y.o.)	38	6,346	98,953	93,767
Secondary (12–16 y.o.)	34	5,682	75,155	70,837
Pre-university (17–21 y.o.)	13	8,542	144,326	84,908

Note: The given dataset statistics correspond to the mobility definition provided in Sec. 4.2.1.

4.1.2 Vertical Mobility Dataset for Estimating Building Floor Height

Table 3 NSE 2016 Vertical Mobility Dataset B: Maximum period of 3 days per subject.

Population	No. schools	No. subjects	No. VD
All	85	20,570	389,031
Primary (6–11 y.o.)	38	6,346	115,465
Secondary (12–16 y.o.)	34	5,682	90,434
Pre-university (17–21 y.o.)	13	8,542	183,132

Note: Trajectories T_i in this dataset contain only upward or downward movement.

Estimating the Height of Elevated Floors: The height of the vertical built environment was estimated by identifying the maximum vertical displacement in a given location (e.g., home, school, or free-range environment) using Vertical Mobility Dataset B (Table 3). In this dataset, displacement is defined as the cumulative distance traversed between consecutive stops without a change in direction—that is, a vertical mobility activity consists of either only upward mobility or only downward mobility. Aside from the difference in what constitutes a trajectory and displacement, the definition of vertical mobility in terms of waiting time ($\triangle t = 15$ minutes) and stop area ($d_{s,v} = 2$ meters) is the same as in the NSE 2016 mobility dataset described in Sec. 4.2.1.

4.2 Data Processing

4.2.1 Mobility Definition

Vertical mobility refers to movement in the built environment's vertical dimension, typically through stairs, escalators, or elevators. A stop in the vertical dimension (S_v) is defined as a location where a user remains within a vertical area of diameter $d_{s,v} = 2$ meters for a waiting time of $\triangle t_v = 5$ minutes. Horizontal mobility refers to movement in the built environment's horizontal plane, i.e., across the urban environment's planar surface. A stop in the horizontal dimension (S_h) is defined as a location where a user remains within a horizontal area of diameter $d_{s,h} = 75$ meters for a waiting time of $\triangle t_h = 15$ minutes.

4.2.2 Derivation of Vertical Mobility

Vertical mobility events were identified using the machine learning framework developed in [37], which is based on a comprehensive understanding of barometric pressure behavior as reviewed in [38]. By leveraging data from the accelerometer, Wi-Fi-based localization, and barometer, this machine learning framework achieved 98% overall accuracy and a 92% F_1 score in classifying vertical mobility events [37]. Additionally, post-processing was applied to remove vertical displacement data around high-pressure fluctuation regions (approximately 8% of the total data) to obtain a conservative estimate. Vertical displacement was calculated based on changes in barometric pressure observed during vertical mobility events. The barometer sensor detects vertical displacement with an accuracy of ± 1 meter. However, due to the relatively low sampling frequency of 1/16 Hz, the minimum identifiable vertical displacement is 2 meters. A barometric pressure change ΔP of 12.01725 Pascals corresponds to a vertical displacement Δr of 1 meter (accurate for $\Delta r < 10$ km above sea level [39]).

4.2.3 Derivation of Horizontal Mobility

Horizontal stops were identified using a standard stop-point detection algorithm provided by the Python library MovingPandas [40]. The algorithm detects a stop when a user remains within an area defined by a diameter $d_{s,h}$ for at least a minimum waiting time $\triangle t_h$. The parameters used in this study are $d_{s,h} = 75$ meters and $\triangle t_h = 15$ minutes. Due to the relatively low spatial resolution, the minimum recognizable horizontal displacement is 20 meters. The horizontal displacement $\Delta r_{h,i}$ is calculated as the sum of the Haversine distances (i.e., great-circle distances) between consecutive location points in a given trajectory $T_{h,i}$.

The choice of the stop area diameter $d_{s,h}$ is informed by the spatial resolution of the Wi-Fi-based localization. The original location data has an accuracy of less than 200 meters for 90% of the data (99% within less than 300 meters). Although the location data has low accuracy, it exhibits relatively high precision, with 91% of consecutive location points having jump lengths of less than 50 meters. However, low accuracy can lead to overestimation of trajectory distances. Therefore, the time series location data was smoothed using a Kalman Filter (KF) employing a nearly-constant velocity (CV) model to reduce noise introduced by low spatial resolution. After smoothing, 98% of

the data has a precision of less than 100 meters, and 96% has less than 50 meters. The KF CV smoother from the MovingPandas library [40] was used with parameter values: (a) process noise uncertainty factor of 0.1 and (b) measurement noise error factor of 50. Results generated with stop diameters of $d_{s,h} = 50$ meters and $d_{s,h} = 100$ meters are provided in the Supplementary Information file (Fig. S2 and Table. S1).

4.2.4 Home-School Detection

The schools' names and locations for each subject were known. Mobility data within a school was identified by locating data points within school zones. School zones were determined using two sources: (a) school zone polygons provided by the Singapore government (data source), and (b) custom square grids drawn around the known school locations with a radius of 150 meters, accounting for systematic deviations in location data (identified visually by plotting all location data from the NSE 2016 dataset).

Home detection involved identifying where users spent most of their nighttime, based on the last detected stop of each day. A custom square grid with a radius of 200 meters centered on the detected home location was used to identify displacements that occurred at or originated from home.

4.2.5 Heavy-Tailed Distribution Fit

The distributions of displacement and waiting time were found to exhibit heavy-tailed behavior and were modeled using one or more (multi-scale) of the following distributions: Power law, Truncated power law, Log-normal, Stretched exponential (also known as the complementary cumulative distribution function (CCDF) of the Weibull distribution), and Exponential distribution. All of these distributions are considered plausible candidates to fit our empirical observations. For each distribution type, the parameters that best fit the empirical data were determined using Maximum Likelihood Estimation (MLE). Once the optimal parameters were found, the candidate distributions were compared to identify the best fit. The log-likelihood and the number of free parameters of each distribution were used to calculate the Akaike Information Criterion (AIC). The best-fitting distribution among the candidates was selected based on the highest Akaike weight (AIC^w), which is calculated from the relative log-likelihoods of the candidate fits.

4.3 Uncertainty and Variability:

All key values are represented as median and median absolute deviation ($\tilde{x} \pm MAD$) or mean and standard deviation ($\mu \pm \sigma$). In appropriate places, $\mu \pm \sigma$ is based on bootstrap values where 100 subjects' data is randomly drawn 1000 times. The following variability tests are conducted and included in the Supplementary Information file. Waiting time variation: $\triangle t_v = 5$, 10, and 15 mins, and $\triangle t_h = 10$, 15, and 20 mins, stop area diameter variation: 50, 75, and 100 m, and displacement definition variation: Trajectory distance vs. straight-line/Origin-Destination OD distance.

References

[1] Wang, C., Hess, D.B.: Role of urban big data in travel behavior research. Transportation research record **2675**(4), 222–233 (2021)

[2] Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G.: Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PloS one **11**(2), 0149222 (2016)

[3] Jiang, S., Ferreira, J., Gonzalez, M.C.: Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. IEEE Transactions on Big Data **3**(2), 208–219 (2017)

[4] Singapore, B.: Building and Construction Authority. (accessed Mar. 24, 2025). https://www1.bca.gov.sg/regulatory-info/lifts-escalators/lift-and-escalator-safety

[5] Department of Statistics, Singapore: Census of Population, 2020: Statistical Release 2: Households, Geographic Distribution, Transport, and Difficulty in Basic Activities. Department of Statistics, Ministry of Trade and Insdustry, Republic of Singapore (2021). https://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr2

[6] Rode, P., Gomes, A., Adeel, M., Sajjad, F., McArthur, J., Alshalfan, S., Schwinger, P., Tunas, D., Lange, C., Montagne, C., et al.: Resource urbanisms: Asia's divergent city models of Kuwait, Abu Dhabi, Singapore and Hong Kong (2017)

[7] Cao, Y., Huang, X.: A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. Remote Sensing of Environment **264**, 112590 (2021)

[8] Komadina, A., Mihajlović, Ž.: Automated 3d urban landscapes visualization using open data sources on the example of the city of Zagreb. KN-Journal of cartography and geographic information **72**(2), 139–152 (2022)

[9] Sun, G., Webster, C., Zhang, X.: Connecting the city: A three-dimensional pedestrian network of Hong Kong. Environment and Planning B: Urban Analytics and City Science **48**(1), 60–75 (2021)

[10] Zhang, L., Chiaradia, A.J.: Walking in the cities without ground, how 3d complex network volumetrics improve analysis. Environment and Planning B: Urban Analytics and City Science **49**(7), 1857–1874 (2022)

[11] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.-L.: Returners and explorers dichotomy in human mobility. Nature Communications **6**(1), 8166 (2015)

[12] Li, J., Tartarini, F., *et al.*: Changes in air quality during the COVID-19 lockdown in Singapore and associations with human mobility trends. Aerosol and Air Quality Research **20**(8), 1748–1758 (2020)

[13] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M.: Human mobility: Models and applications. Physics Reports **734**, 1–74 (2018)

[14] Poletto, C., Tizzoni, M., Colizza, V.: Human mobility and time spent at destination: impact on spatial epidemic spreading. Journal of Theoretical Biology **338**, 41–58 (2013)

[15] Manivannan, A., Chin, W.C.B., Gopalakrishnan, S., Wong, D.K., Schroepfer, T., Bouffanais, R.: Effects of the interplay between topology and function of an integrated urban development on patterns of user movement. Scientific Reports **14**(1), 7021 (2024)

[16] Webb, B., White, J.T.: Planning and the high-rise neighbourhood: Debates on vertical cities. Urban Planning **7**(4), 208–212 (2022)

[17] Bruyns, G.J., Higgins, C.D., Nel, D.H.: Urban volumetrics: From vertical to volumetric urbanisation and its extensions to empirical morphological analysis. Urban Studies **58**(5), 922–940 (2021)

[18] Maloutas, T.: The role of vertical segregation in urban social processes. Nature Cities **1**(3), 185–193 (2024)

[19] Nethercote, M.: Melbourne's vertical expansion and the political economies of high-rise residential development. Urban Studies **56**(16), 3394–3414 (2019)

[20] Graham, S.: Luxified skies: How vertical urban housing became an elite preserve. City **19**(5), 618–645 (2015)

[21] Xu, Y., Belyi, A., Bojic, I., Ratti, C.: Human mobility and socioeconomic status: Analysis of Singapore and Boston. Computers, Environment and Urban Systems **72**, 51–67 (2018)

[22] Frolking, S., Mahtta, R., Milliman, T., Esch, T., Seto, K.C.: Global urban structural growth shows a profound shift from spreading out to building up. Nature Cities, 1–12 (2024)

[23] De Almeida, A., Hirzel, S., Patrão, C., Fong, J., Dütschke, E.: Energy-efficient elevators and escalators in europe: An analysis of energy efficiency potentials and policy measures. Energy and Buildings **47**, 151–158 (2012)

[24] Tukia, T., Uimonen, S., Siikonen, M.-L., Donghi, C., Lehtonen, M.: Modeling the aggregated power consumption of elevators–the new york city case study. Applied

Energy **251**, 113356 (2019)

[25] Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A.: Multi-scale spatio-temporal analysis of human mobility. PloS one **12**(2), 0171686 (2017)

[26] Liu, T., Yang, Z., Zhao, Y., Wu, C., Zhou, Z., Liu, Y.: Temporal understanding of human mobility: A multi-time scale analysis. PloS one **13**(11), 0207697 (2018)

[27] Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S.: Explaining the power-law distribution of human mobility through transportationmodality decomposition. Scientific Reports **5**(1), 1–7 (2015)

[28] Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C.: Geo-located Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science **41**(3), 260–271 (2014)

[29] Yan, X.-Y., Han, X.-P., Wang, B.-H., Zhou, T.: Diversity of individual mobility patterns and emergence of aggregated scaling laws. Scientific Reports **3**(1), 2678 (2013)

[30] Liu, Y., Kang, C., Gao, S., Xiao, Y., Tian, Y.: Understanding intra-urban trip patterns from taxi trajectory data. Journal of Geographical Systems **14**(4), 463–483 (2012)

[31] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S.J., Chong, S.: On the levy-walk nature of human mobility. IEEE/ACM Transactions on Networking **19**(3), 630–643 (2011)

[32] Song, C., Koren, T., Wang, P., Barabási, A.-L.: Modelling the scaling properties of human mobility. Nature Physics **6**(10), 818–823 (2010)

[33] Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)

[34] Pappalardo, L., Manley, E., Sekara, V., Alessandretti, L.: Future directions in human mobility science. Nature Computational Science **3**(7), 588–600 (2023)

[35] Alessandretti, L., Natera Orozco, L.G., Saberi, M., Szell, M., Battiston, F.: Multimodal urban mobility and multilayer transport networks. Environment and Planning B: Urban Analytics and City Science **50**(8), 2038–2070 (2023)

[36] Wilhelm, E., Siby, S., Zhou, Y., Ashok, X.J.S., Jayasuriya, M., Foong, S., Kee, J., Wood, K.L., Tippenhauer, N.O.: Wearable environmental sensors and infrastructure for mobile large-scale urban deployment. IEEE Sensors Journal **16**(22), 8111–8123 (2016)

[37] Manivannan, A., Willemse, E.J., Chin, W.C.B., Zhou, Y., Tunçer, B., Barrat, A., Bouffanais, R., *et al.*: A framework for the identification of human vertical

displacement activity based on multi-sensor data. IEEE Sensors Journal **22**(8), 8011–8029 (2022)

[38] Manivannan, A., Chin, W.C.B., Barrat, A., Bouffanais, R.: On the challenges and potential of using barometric sensors to track human activity. Sensors **20**(23), 6786 (2020)

[39] Young, D.F., Munson, B.R., Okiishi, T.H., Huebsch, W.W.: A Brief Introduction to Fluid Mechanics. John Wiley & Sons, New Jersey, USA (2010)

[40] Graser, A.: MovingPandas: Efficient Structures for Movement Data in Python. GI_Forum – Journal of Geographic Information Science **7**(1), 54–68 (2019) https://doi.org/10.1553/giscience2019_01_s54 . Accessed 2023-04-19

Supplmentary Information: Vertical Movement in Cities: Uncovering the Hidden Dimension of Urban Mobility

Ajaykumar Manivannan[1], Elias J. Willemse[2], Wei Chien Benny Chin[3], Bige Tunçer[4], Alain Barrat[5], Roland Bouffanais[1,6*]

[1*]Department of Mechanical Engineering, University of Ottawa, Ottawa, K1N6N5, Ontario, Canada.
[2]Waste Labs Pte. Ltd., 059915,Singapore.
[3]Department of Geography, National University of Singapore, 117568, Singapore.
[4]Information Systems in the Built Environment, The Eindhoven University of Technology, City, 5612 AZ, North Brabant, Netherlands.
[5]Center for Theoretical Physics, CNRS and Aix Marseille University, Marseille, 13009, Provence-Alpes-Côte d'Azur, France.
[6]Department of Computer Science & Global Studies Institute, University of Geneva, Genève CH-1211, Switzerland.

*Corresponding author(s). E-mail(s): roland.bouffanais@unige.ch;

1 Daily time coverage per subject

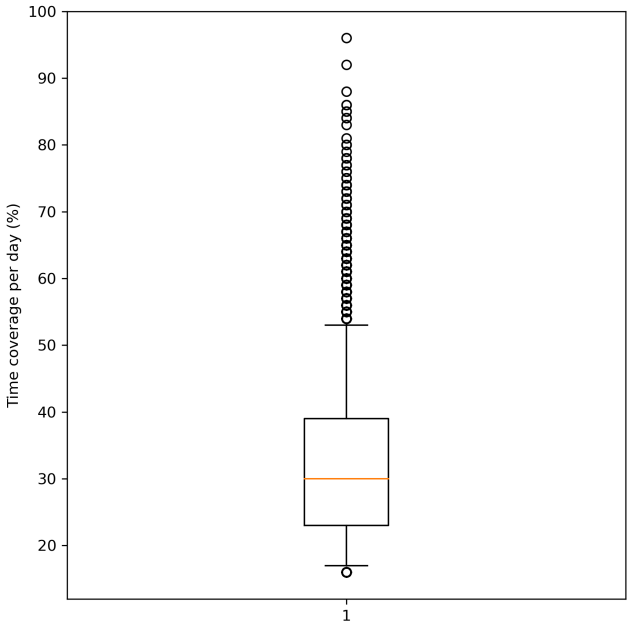


Fig. S1 Daily time coverage per subject. Here, time coverage is when active body movements were detected. The minimum time coverage used to filter a given day for a subject is 4 hours (16.7%), and the median time coverage among the population is 7.2 hours (30%). Whiskers of the boxplot represent the 5^{th} and 95^{th} percentile.

2 Choosing parameters that define mobility:

The choice of waiting time $\triangle t$ should reasonably reflect the mobility behavior of humans. For example, we define displacement as the distance a user travels when motivated to move from one place to another. While assessing what motivates an individual is complex, the stop point should ideally ignore re-fuelling, transit changes, or traffic stops that are not directly relevant to what makes a user move from one place to another.

In introducing mobility in the vertical dimension, we must consider an appropriate waiting time $\triangle t_v$ different from the factors influencing our choice for $\triangle t_h$. For example, some key factors to consider are (a) the spatiotemporal scale of vertical mobility is much smaller, (b) transit time between vertical transportation modes such as a lift is also typically smaller than the transit time between two buses or trains, and finally (c) motivated destination within a building or a development complex is also plenty and diverse than in the horizontal dimension at the urban scale [1]. For these reasons, the $\triangle t_v$ is chosen to be smaller than $\triangle t_h$. Results generated with equal wait times ($\triangle t_v = \triangle t_h = 15$ minutes) are shown in subsequent sections. The choice of stop radius is informed by the minimum spatial and temporal resolution available in the NSE 2016 data set (discussed later in the section).

Our mobility data set does not sufficiently represent movement within $\pm d_s/2$ meters of a stop, as is typically expected by choice of d_s. Results generated with different waiting times ($\triangle t_v = 10$ and 15 minutes and $\triangle t_h = 10$ and 20 minutes) are shown in subsequent sections.

3 Effect of variation of waiting time, stop area diameter, and displacement type to define mobility

Sensitivity of $P(x)$: The distribution of displacement, waiting time, and mobility duration have a similar form for different parameters that define mobility based on waiting time (10, 15, and 20 minutes for HM and 5, 10, and 15 minutes for VM), stop area diameter (50, 75, or 100 meters for HM), or distance type (trajectory/cumulative distance or origin-destination/straight line distance).

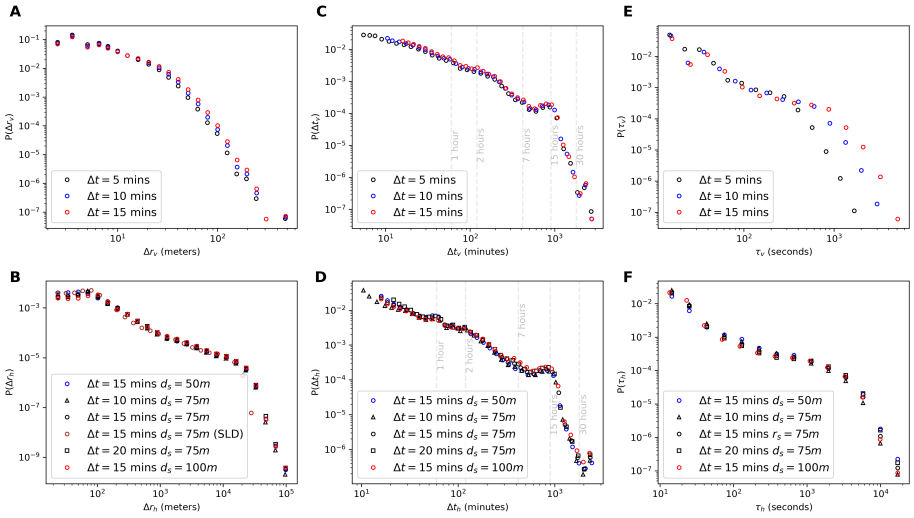


Fig. S2 Sensitivity of distribution of mobility properties to waiting time, stop area diameter, and displacement definition. All graphs are in a log-log plot with logarithmic binning. The scatter points denote the mobility properties' probability density function $P(x)$ for $x > x_{min}$. $P(x)$ for **A & B,** displacement $\triangle r$, **C & D,** waiting time $\triangle t$, and **E & F,** mobility duration τ for respective mobility dimensions.

Sensitivity of daily mobility statistics: Results are overall consistent for metrics showing the magnitude of displacement while different for mobility duration and waiting time for different waiting times and stop area diameter used to define mobility. This is, however, expected by the definition of the parameters. For example, increasing waiting times will absorb more displacements into a single trip, thus improving overall waiting times. Hence, appropriate parameters to define mobility should be based on practical considerations such as excluding traffic or fuel stops. The difference in daily mobility statistics is also more pronounced by age group, as the population of 17+ subjects is expected to travel more to reach the sparsely located schools. In contrast, the younger populations go to school within or closer neighborhoods, thus commuting relatively smaller distances.

Type	Population	$N_{subjects}$	N_{days}	$N_{\triangle r}$	$\triangle r^{num}$	$\triangle r^{num/hour}$	$\triangle r^{vol}$ (m)	$\triangle r^{max}$ (m)	τ^{vol} (mins)	$\triangle t^{vol}$ (mins)
VM $\triangle t$ = 5 mins	all	20,571	43,032	318,434	7± 2	0.78± 0.22	77.56± 32.7	22.8± 8.32	7 ±4.15	61.53± 33.16
VM $\triangle t$ = 5 mins	17-21 y.o.	8,543	15,942	144,326	9± 2	0.9± 0.22	102.35± 32.95	24.71± 6.57	11± 5	44.28± 18.55
VM $\triangle t$ = 5 mins	6-11 y.o.	6,346	14,936	98,953	6± 2	0.74± 0.19	64.32± 27.71	20.89± 9.4	5.1± 2.9	84.68± 45.09
VM $\triangle t$ = 5 mins	12-16 y.o.	5,682	12,154	75,155	6± 2	0.66± 0.21	62.33± 27.54	21.55± 9.07	5.2± 3.2	77.14± 48.69
VM $\triangle t$ = 10 mins	all	20,571	43,032	281,289	6± 2	0.7± 0.19	77.56± 32.7	25.13± 9.4	11.5± 8	79.85± 39.82
VM $\triangle t$ = 10 mins	17-21 y.o.	8,543	15,942	123,713	7± 2	0.78± 0.18	102.35± 32.95	27.63± 7.82	19± 9.7	58.69± 23.16
VM $\triangle t$ = 10 mins	6-11 y.o.	6,346	149,36	901,89	6± 2	0.68± 0.17	64.32± 27.71	22.8± 10.32	8.3± 5.6	101.35± 47.9
VM $\triangle t$ = 10 mins	12-16 y.o.	5,682	121,54	67,387	5 ±2	0.61± 0.18	62.33± 27.54	23.72± 10.15	8± 5.9	99.87± 59.22
VM $\triangle t$ = 15 mins	all	20,571	43,032	256,508	6± 2	0.64± 0.17	77.56± 32.7	27.13 ±10.4	17.3 ±12.3	95.22± 44.57
VM $\triangle t$ = 15 mins	17-21 y.o.	8,543	15,942	110,931	7± 2	0.7± 0.15	102.35± 32.95	30.21± 8.9	28.2± 14	71.7± 27.2
VM $\triangle t$ = 15 mins	6-11 y.o.	6,346	149,36	83,730	5± 2	0.63± 0.16	64.32± 27.71	24.13± 10.73	12.9 ±9.2	117.1 ±52.45
VM $\triangle t$ = 15 mins	12-16 y.o.	5,682	121,54	61,847	5± 2	0.56± 0.16	62.33± 27.54	25.3± 11.07	12.3± 9.8	120.76± 67.76
HM $\triangle t$ = 15 mins d_s = 50 m	all	20,571	43,165	267,270	6± 2	0.66± 0.17	11260.07± 8595.86	6102.75 ±4853.93	117.3± 49.9	80.87± 31.99
HM $\triangle t$ = 15 mins d_s = 50 m	17-21 y.o.	8,543	15,953	88,713	5± 1	0.6 ±0.14	26,367.81± 12,453.57	14,201.67± 6,294.13	159.2± 45.9	93.64± 34.51
HM $\triangle t$ = 15 mins d_s = 50 m	6-11 y.o.	6,346	15,016	102,136	7± 2	0.73± 0.17	4,320.37± 2,896.66	2,111.43± 1,537.98	82.8± 37.9	71.06± 25.44
HM $\triangle t$ = 15 mins d_s = 50 m	12-16 y.o.	5,682	12,196	76,421	6± 2	0.66± 0.19	8,371.11± 4,836.1	4,436.05± 2,643.52	102.7± 40.1	74.17± 30.31
HM $\triangle t$ = 10 mins d_s = 75 m	all	20,569	43,136	281,920	6 ±2	0.69± 0.19	11,153.47 ± 8,667.51	5,915.54± 4,760.45	85.5± 42.9	76.35± 35.25
HM $\triangle t$ = 10 mins d_s = 75 m	17-21 y.o.	8,543	15,951	96,378	6 ±2	0.64± 0.16	26,399.03± 12,505.39	13,947.06± 6,233.85	131.1± 39.5	86.02± 37.06
HM $\triangle t$ = 10 mins d_s = 75 m	6-11 y.o.	6,346	14,996	105,933	7± 2	0.76± 0.19	4,129.44± 2,871.81	2,005.19± 1,497.32	50.5± 26.3	70.15± 31.03
HM $\triangle t$ = 10 mins d_s = 75 m	12-16 y.o.	5,680	121,89	79,609	6± 2	0.68± 0.22	8,244.85± 4,866.42	4,283.67± 2,580.16	74.1± 30.6	70.89± 33.42
HM $\triangle t$ = 15 mins d_s = 75 m	all	20,571	43,128	249,512	6± 2	0.63± 0.17	11,177.78± 8,667.72	6,053.37± 4,832.51	95.7± 45.2	91.98± 39.62
HM $\triangle t$ = 15 mins d_s = 75 m	17-21 y.o.	8,543	15,950	84,908	5± 1	0.58 ±0.14	26,372.59± 12,456.7	14,138.96± 6,288.94	141.4± 41.3	101.88 ±40.06
HM $\triangle t$ = 15 mins d_s = 75 m	6-11 y.o.	6,346	14,988	93,767	6± 2	0.69± 0.18	4,130.36± 2,861.74	2,055.24± 1,518.35	59.2± 29.6	85.36± 35.92
HM $\triangle t$ = 15 mins d_s = 75 m	12-16 y.o.	5,682	12,190	70,837	6± 2	0.62± 0.2	8,238.15± 4,851.42	4,393.96± 2,627.64	83.3± 33.1	87.25± 41.5
HM $\triangle t$ = 20 mins d_s = 75 m	all	20,571	43,121	228,624	5± 1	0.58± 0.16	11,183.61± 8,657.31	6,129.12± 4,881.3	102.6± 47.7	104.23 45.16
HM $\triangle t$ = 20 mins d_s = 75 m	17-21 y.o.	8,543	15,948	79,257	5± 1	0.55± 0.13	26,331.25± 12,457.08	14,257.61± 6,308.24	149.1± 43.3	110.01± 42.69
HM $\triangle t$ = 20 mins d_s = 75 m	6-11 y.o.	6,346	14,983	84,853	5± 2	0.63± 0.16	4,134.02± 2,866.71	2,106.74 ±1,542.57	65.8± 32.6	96.84 ±40.93
HM $\triangle t$ = 20 mins d_s = 75 m	12-16 y.o.	5682	12,190	64,514	5± 2	0.57± 0.18	8,223.01± 4,831.17	4,447.68± 2,643.75	90.6± 36.6	101.85± 49.93
HM $\triangle t$ = 15 mins d_s = 100 m	all	20,571	43,104	223,018	5± 2	0.58± 0.17	11,054.97± 8,704.65	6,001.74± 4,817.71	83.2± 42.4	110.83 ±54.51
HM $\triangle t$ = 15 mins d_s = 100 m	17-21 y.o.	8,543	15948	80399	5± 1	0.56± 0.14	26,263.2± 12,459.56	14,084.35± 6,274.22	131± 39.2	110.63± 46.8
HM $\triangle t$ = 15 mins d_s = 100 m	6-11 y.o.	6,346	14,967	79,930	5± 2	0.6± 0.18	3,966.01± 2,835.16	2,023.46 ±1,514.39	47.3± 25.2	112.38± 57.56
HM $\triangle t$ = 15 mins d_s = 100 m	12-16 y.o.	5,682	12,189	62,689	5± 2	0.56± 0.19	8,112.84± 4,859.3	4,360.5± 2,628.11	72± 29.7	109.35± 60.55

Table S1 Daily mobility statistics for different waiting times ($\triangle t$) and stop area diameter (d_s). HM: Horizontal mobility, VM: Vertical Mobility. For each cell value with format $\tilde{x}\pm$ MAD: $\tilde{x}$ represents median value, and MAD represents median absolute deviation, i.e., $\widetilde{|x_i - \tilde{x}|}$; The number ($num$) and volume ($vol$) of properties are the count and sum of the given property value x.

4 Inferred residential height for different age groups

Families with young children are just as likely to live across all floor levels as families with adult children (17+).

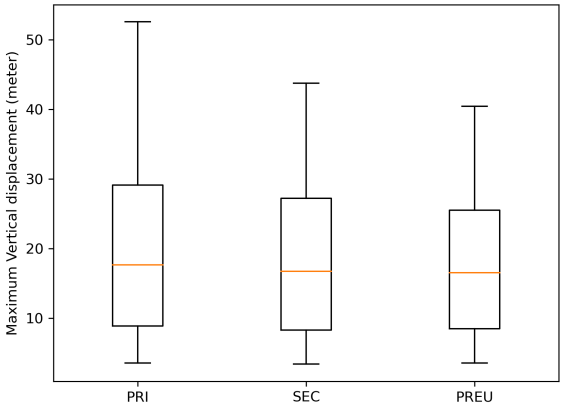


Fig. S3 Inferred residential height for different age groups. PRI: 6 to 11; SEC: 12 to 16; PREU: 17 to 21. Whiskers of the boxplot represent the 5^{th} and 95^{th} percentile.

5 Heavy-tailed Distribution fit:

The distribution of the displacement and the waiting time has been found to have one or more (multi-scale) of the following heavy-tailed distributions – Power law, Truncated power law, Log-normal, Stretched exponential (also known as CCDF of Weibull distribution), and Exponential distribution. All of these distributions are considered plausible candidates to fit our empirical observations. For each distribution type, the respective parameters that best fit the empirical data are found by Maximum Likelihood Estimate (MLE). Once the optimal parameters of the distribution are found, the distributions are compared with each other to find the best fit. The log-likelihood and the number of free parameters for the distributions are used to calculate the Akaike Information Criterion (AIC). The best fit among the candidates is the one with the maximum AIC weight (AIC^w) calculated from the relative log-likelihood of the candidate fits.

AIC calculations:

For a distribution model m,

$$AIC_m = -2 * LogL_m + 2 * V_m + \frac{2 * V_m(V_m + 1)}{n - V_m - 1}, \tag{1}$$

where L_m is the maximum likelihood for the candidate model m, V_m is the number of free parameters of the model, and n is the number of sample data.

$$AIC^w = \frac{e^{-0.5(AIC_m - AIC_{min})}}{\sum_{k=1}^{K} e^{-0.5(AIC_k - AIC_{min})}}, \tag{2}$$

where K is the number of candidate models and AIC_{min} is the minimal AIC score found among the candidates.

Kolmogorov-Smirnov (KS) statistics measure the distance between the synthetic or theoretical fit and the empirical data. It is used to find the x_{min} and x_{max} values. It is given by,

$$KS = max(|F - P|), \tag{3}$$

where F is the cumulative distribution of the best fit, and P is the cumulative distribution of the empirical or synthetic data.

Robustness of the fit: By bootstrapping data 1000 times for 100 subject's data, we find if the candidate model that fits the overall empirical data is also the best fit for its sub-samples. We also measure the uncertainty of the fit parameters and present it through its mean and standard deviation, i.e., $\mu \pm \sigma$.

The best fit for all of the mobility properties is either the stretched exponential or truncated power -law (Table S2). This is robust across the sub-samples where the best candidate and its parameters mostly remain the same (Table S3).

Table S2 Heavy-tailed distribution fit.

Metric	x_{min}	N	p_{PL}	$\mathcal{R}_{TPL}$	$\mathcal{R}_{LN}$	$\mathcal{R}_{SE}$	$\mathcal{R}_{E}$	KS_{BF}	AIC^w_{BF}
$\triangle r_v$	2 m	316,686	0	-42,426	-41,666	**-44,471**	-36,838	0.0276	1.0
$\triangle t_v$	5 min	297,877	0	-59,466	-57,181	**-60,625**	12,341	0.0224	1.0
τ_v	13 sec	318,380	0	**-5,662**	-2,804	-2,328	100,800	0.107	1.0
$\triangle r_h$	40 m	234,822	0	**-22,210**	-12,889	-15,311	187,103	0.0589	1.0
$\triangle t_h$	15 min	228,941	0	**-29,853**	-27,052	-29,000	8,979	0.0283	1.0
τ_h	10 sec	249,510	0	**-57,581**	-32,092	-44,358	71,147	0.130	1.0

N: Number of data points where $x \in x_{min}, \infty$; PL: Power Law; TPL: Truncated Power Law; LN: Lognormal; SE: Stretched exponential; E: Exponential; BF: Best Fit; p_{PL}: p-value for Power Law. It is defined as the fraction of the synthetic distances larger than the empirical distance. Distance is given by Kolmogorov-Smirnov statistics. $\mathcal{R}$: Loglikelihood ratio between Power law and the given distributions. A negative value indicates that the given candidate model is better than the power-law, and the magnitude of the value indicates its strength in that direction. The significance of the sign and magnitude is given by the p-value (p-value$< 10^{-6}$ for all $\mathcal{R}$ values). The best fit for each metric is highlighted in bold, and its respective $AIC^w = 1$.

Table S3 Robustness of best distribution fit and its parameters.

Metric	BF	BF (bootstrap)	BF parameter (bootstrap)
$\triangle r_v$	SE	SE (93.9%), TPL (6.4%)	$\beta = 0.72 \pm 0.06$
$\triangle t_v$	SE	SE (83.4%), TPL (16.6%)	$\beta = 0.485 \pm 0.019$
τ_v	TPL	TPL (100%)	$\alpha = 1.68 \pm 0.05$, $\lambda = 0.0013 \pm 0.0002$
$\triangle r_h$	TPL	TPL (100%)	$\alpha = 1.15 \pm 0.03$, $\lambda = 3.228 * 10^{-5} \pm 4.139 * 10^{-6}$
$\triangle t_h$	TPL	TPL (94.2%), SE (5.8%)	$\alpha = 1.015 \pm 0.02$, $\lambda = 0.00183 \pm 0.0001$
τ_h	TPL	TPL (100%)	$\alpha = 1.0 \pm 0.0007$, $\lambda = 0.000149 \pm 0.00001$

The robustness of the best fit is tested by a bootstrapping method where data from 100 subjects are randomly drawn 1000 times, and the resulting best-fit parameters are given by mean ± standard deviation. BF: Best Fit; AIC^w: Akaike Information Criterion weight; KS: Kolmogorov-Smirnov statistics between empirical data and theoretical fit

6 Specific energy consumption of vertical and horizontal transportation

Specific energy consumption E_{spc} is the standardized energy consumed per passenger per unit distance. Its unit is given by Joules/(passenger*meter) or J/pax.m. Table S4 shows E_{spc} for different transportation modes. The main article reports E_{spc} for vertical and horizontal transportation. It is based on literature-reported values in cities comparable to Singapore, such as Hong Kong, London, and New York. Table S4 also lists E_{spc} as a reference only for other countries in the US or Canada whose overall urban and transportation infrastructure can be vastly different from cities like Singapore.

$E_{spc,v}$ **for vertical transportation:** Based on monitoring 71 elevators in four countries in Europe, a comprehensive study by [2] reports that hydraulic ($<$ 6 floors), geared (6 to 25 floors), and gearless ($>$ 26 floors) elevators consume 14.1, 5.5, and 3 mWh/Kg.m of energy respectively during running. Taking a weighted average of hydraulic (20 %), geared (60 %), and gearless (20 %), we calculate $E_{running} \sim 1220.4$ J/pax.m. It is also reported that the standby energy is 50% to 65% (on average) of the total energy consumed by an elevator [2]. Assuming at least $E_{standby} \sim 1 * E_{running}$, we calculate total $E_{spc} = E_{standby} + E_{running} = 2440$ J/pax.m. This is on the lower side of the estimates reported in Table S4, including considering escalators ($E_{spc} = 4000$ J/pax.m). In this article, we use the relatively lower value of $E_{spc,v} \sim 2440$ J/pax.m for conservative comparison.

$E_{spc,h}$ **for horizontal transportation:** We report here two estimates from literature—(1) Rode et al. [3] estimates a weighted average of specific energy consumed in Hong Kong to be $E_{spc,h} \sim 500$ J/pax.m. This is based on the following modes: car (8 %): 2300 J/pax.m, bus (45 %): 500 J/pax.m, and train (36 %): 200 J/pax.m. (2) Based on values reported in comparable cities from multiple sources, we estimate $E_{spc,h} \sim 652$ J/pax.m by keeping the same weights for modes as above. This is based on the following modes: car (London): 1050 J/pax.m [4], bus (Singapore): 800 J/pax.m [5], and train (Hong Kong): 291 J/pax.m [6]. In this article, we use the relatively higher value of $E_{spc,h} \sim 652$ J/pax.m for conservative comparison.

Table S4 Specific energy consumption for different modes of transport

Mode	Desc.	EC (literature)	Capacity (pax)	Load factor (pax/car)	EC (J/pax-m)
Train [6]	Toronto, Canada	1.19 MJ/pax-km	NA	NA	1190
Train [7]	US DoE Commuter rails	1.050 MJ/pax-km	NA	33.6	1050
Train [7]	US DoE Transit rails	0.553 MJ/pax-km	NA	23.5	553
Train [6]	Hong Kong Urban rail	0.081 kWh/pax-km	312.5	59.7	**291**
Bus [7]	US DoE transit buses and trolleys	2.989 MJ/pax-km	NA	7.7	2989
Bus [6]	Hong Kong Bus	1.056 MJ/pax-km	150	28.7	1056
Bus [5]	Singapore diesel—single-decker (estimated)	4 kWh/km (average per bus)	90	18 [1]	**800**
Bus [5]	Singapore electric—all vehicles types (estimated)	1.9 kWh/km (average per bus)	90-185	18.8	364
Car [7]	US DoE	1.86 MJ/pax-km	NA	1.5	1861
Car - ICE [4]	Average of 14 vehicles in UK	1.68 MJ/km	NA	1.6 $*$	**1050**
Car - EV [4]	Average of 16 vehicles in UK	0.615 MJ/km	NA	1.6	384
Elevator [8]	Average of 4 buildings (Malaysia)	$E_{running} \sim$ 10 kW at 1 m/s is 10,000 J/m	NA	2.5 $*$	4000
Escalator [9]	Train station	12 kW for 6.3m at 0.5 m/s is 24,000 J/m	$\sim$ 30	6 $*$	4000
Elevator [9]	Train station	7.4 kW for 6.3m at 1 m/s is 7400 J/m	13	2.6 $*$	2846
Elevator [2]	Europe (weighted estimate)	$E_{running} \sim$ 4.52 mWh/kg.m + $E_{standby} \sim 1 * E_{running}$	NA	NA	**2440**
Elevator [10]	Europe (estimated)	$E_{running}$ (mean) $\sim$ 2 mWh/kg.m + $E_{standby} \sim 1 * E_{spc}$	NA	NA	1080

Note: Modes highlighted in bold are used to calculate E_{spc}. 1 kWh = $3.6 * 10^6$ J; 1 kW = 1 kJ/s; 1 kW = (1/velocity) kJ/m; 1 mWh/kg.m = 270 J/m.pax (assuming pax weighs 75 kg) EC: Energy Consumption; CFP: Carbon Foot Print; EV: Electric Vehicle; ICE: Internal Combustion Engine.

$*$ assuming load factor = 20% of maximum capacity

References

[1] Manivannan, A., Chin, W.C.B., Gopalakrishnan, S., Wong, D.K., Schroepfer, T., Bouffanais, R.: Effects of the interplay between topology and function of an integrated urban development on patterns of user movement. Scientific Reports **14**(1), 7021 (2024)

[2] De Almeida, A., Patrão, C., Fong, J., Nunes, U., Araújo, R., Nunes, U., Rivet, L., Lindegger, U., Nanetti, M., Cariani, W., et al.: E4-energy efficient elevators and escalators,". Technical report, SR-University of Coimbra (Portugal) (2009)

[3] Rode, P., Gomes, A., Adeel, M., Sajjad, F., McArthur, J., Alshalfan, S., Schwinger, P., Tunas, D., Lange, C., Montagne, C., et al.: Resource urbanisms: Asia's divergent city models of Kuwait, Abu Dhabi, Singapore and Hong Kong (2017)

[4] Howey, D.A., Martinez-Botas, R., Cussons, B., Lytton, L.: Comparative measurements of the energy consumption of 51 electric, hybrid and internal combustion engine vehicles. Transportation Research Part D: Transport and Environment **16**(6), 459–464 (2011)

[5] Gallet, M., Massier, T., Hamacher, T.: Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Applied Energy **230**, 344–356 (2018)

[6] To, W., Lee, P.K., Billy, T.: Sustainability assessment of an urban rail system–the case of hong kong. Journal of cleaner production **253**, 119961 (2020)

[7] Davis, S.C., Boundy, R.G.: Transportation energy data book: Edition 39. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States) (2021)

[8] Ang, J.H., Yusup, Y., Zaki, S.A., Salehabadi, A., Ahmad, M.I.: Comprehensive energy consumption of elevator systems based on hybrid approach of measurement and calculation in low-and high-rise buildings of tropical climate towards energy efficiency. Sustainability **14**(8), 4779 (2022)

[9] Casals, M., Gangolells, M., Forcada, N., Macarulla, M., Giretti, A.: A breakdown of energy consumption in an underground station. Energy and buildings **78**, 89–97 (2014)

[10] De Almeida, A., Hirzel, S., Patrão, C., Fong, J., Dütschke, E.: Energy-efficient elevators and escalators in europe: An analysis of energy efficiency potentials and policy measures. Energy and Buildings **47**, 151–158 (2012)

Part IV

What influences movement in an integrated development? A Case Study

7

Effects of the interplay between topology and function of an integrated urban development on patterns of user movement

This chapter includes a peer-reviewed paper published in Scientific Reports, which addresses the fourth research question (R4) posed in Chapter 1, focusing on a case study that aims to understand the factors that influence user movement (vertical and horizontal) in an integrated development.

Publication details:

- Manivannan, A., Chin, W. C. B., Gopalakrishnan, S., Wong, D. K., Schroepfer, T., & Bouffanais, R. (2024). Effects of the interplay between topology and function of an integrated urban development on patterns of user movement. Scientific Reports, 14(1), 7021.

Related publications, conference presentations, and conference proceedings (not attached or described in the thesis):

- Conference presentation: A. Manivannan, S. Gopalakrishnan, D. Wong, R. Bouffanais, and T. Schroepfer, Informed design of future integrated developments using complexity science, Presented by Manivannan, A. to the Conference on Complex Systems (CCS 2020), online, 4-11 December 2020, 2020. DOI: `https://doi.org/10.5281/zenodo.4419178`. [Online]. Available: `https://www.youtube.com/watch?v=9ed74ypA20Y`.

- Conference presentation: A. Manivannan, S. Gopalakrishnan, R. Bouffanais, and T. Schroepfer, Vertical cities: Complex emergent patterns of movement and space use in high-density urban contexts, Presented by Manivannan, A., and Gopalakrishnan, S., to the MacArthur Workshop on Urban Modelling & Complexity Science, Centre for Advanced Spatial Analysis, UCL, UK. 16-20 September 2019, 2019. [Online]. Available: https://www.ucl.ac.uk/bartlett/casa/events/2019/sep/macarthur-workshop-urban-modelling-complexity-science.

- Journal paper: Gopalakrishnan, S., Wong, D., Chin, B., Srikanth, A. D., Manivannan, A., Bouffanais, R., & Schroepfer, T. (2023). Vertical Cities: Emergent Patterns of Movement and Space Use in Dense Vertically Integrated Urban Built Environments. International Journal on Smart and Sustainable Cities, 1(01), 2340005.

- Conference proceeding: Gopalakrishnan, S., Wong, D., Manivannan, A., Chin, B., Boufannais, R., & Schroepfer, T. (2022, February). User-driven Emergent Patterns of Movement and Landscape Space Use in Vertically Integrated Urban Environments. In 14th Ecocity World Summit 2022.

Summary: This paper presents a data-driven framework for analyzing human movement patterns within an integrated urban development, using the Kampung Admiralty complex in Singapore as a case study. One of the key contributions is the application of network-theoretic methods to quantify the relationship between building topology and user movement, revealing a moderate correlation of about 40% between spatial network features and movement flows. This highlights that while spatial connectivity influences movement patterns, other factors, such as function and user preferences, also play significant roles. The study enhances our understanding of how users navigate multi-level, mixed-use environments, especially the complex interplay between horizontal and vertical mobility across different functions, such as residential, commercial, and public spaces.

Another key contribution of the research is its exploration of how spatial function influences user movement, particularly in high-density vertical developments. By aggregating movement data and analyzing flows between floors and functional spaces, the study uncovers that certain areas, such as commercial zones and social spaces, attract higher movement due to their connectivity and functional diversity. The general framework developed in this study provides valuable insights for urban designers and planners, helping to optimize circulation in vertically integrated developments and improve the overall user experience in such high-density urban environments. In addition, this framework can also be readily applied to other urban systems.

scientific reports

Check for updates

OPEN

Effects of the interplay between topology and function of an integrated urban development on patterns of user movement

Ajaykumar Manivannan[1,6], Wei Chien Benny Chin[2,6], Srilalitha Gopalakrishnan[3], Daniel K. H. Wong[4], Thomas Schroepfer[3,4] & Roland Bouffanais[5]✉

With the advent of distributed multi-sensory networks of devices, vast troves of real-time data can be gathered about our interactions with the built environment. These rich data sets can be mined to achieve improved and informed data-driven designs of buildings, neighborhoods, and potentially entire cities. Among those, integrated developments have the peculiarity of combining multiple functions within a compact space and, as such, behave as microcosms of a city that can help address the problem of urban sprawl and density. However, a general lack of data and framework about integrated developments hinders our ability to test design hypotheses about the complex interplay between heterogeneity in both space and function. Here, we apply a data-driven approach to analyze the joint influence of topology and function on user movement within a state-of-the-art integrated development in Singapore. Specifically, we leverage the network representation of the building and use movement data collected from 51 individuals over a month. We show evidence of correlation (40%) between the spatial network features and human movement at the building level. We are also able to quantify the relationship between the functional and spatial components of the integrated development through user movement. Previous studies have shown a 60% or higher correlation between the topology and human movement at the city or country scales. Our moderate correlation, therefore, implies that more factors influencing user movement are at play. The heterogeneity in the spatial function introduced trips with diverse origins and destinations. A further data-driven analysis integrating origins and destinations reveals both qualitative and quantitative means of studying the relationship between the built environment and the processes that take place in them.

The future of human settlements is projected to be urban with the ratio of the world population residing in cities expected to increase from 55%, as of today, to 68% in 2050[1]. Urban areas are responding to this unabated trend in mostly two non-mutually-exclusive ways: (1) with a horizontal growth according to the well-known urban sprawl model, and (2) with an increase in density that usually leads to vertical growth, typically when faced with land scarcity and/or strained transportation infrastructure.

As our understanding of cities and their relationship to the socio-economic parameters concerning the human population improves[2,3], better city planning is not only expected, but is an imperative. As the impact of climate change is more visible and pressing, cities are also forced to take their environmental impact into account. Hence, future cities aim to focus on sustainable development along these three verticals—economics, social, and environment. One of the stated seventeen goals of the United Nations (UN) Sustainable Development Goals (SDG) for the year 2030 calls for cities to be inclusive, safe, resilient, and sustainable[4].

[1]Department of Mechanical Engineering, University of Ottawa, Ottawa K1N 6N5, Canada. [2]Department of Geography, National University of Singapore, 117568 Singapore, Singapore. [3]Future Cities Laboratory Global, Singapore-ETH Centre, 138602 Singapore, Singapore. [4]Architecture and Sustainable Design, Singapore University of Technology and Design, 487372 Singapore, Singapore. [5]Department of Computer Science & Global Studies Institute, University of Geneva, 1211 Geneva, Switzerland. [6]These authors contributed equally: Ajaykumar Manivannan and Wei Chien Benny Chin. ✉email: roland.bouffanais@unige.ch

As factors like social welfare (community happiness, health, inclusivity) and environmental sustainability (waste management, energy consumption, environmental impact) are brought to the forefront, future cities look for an integrated solution to these challenges. Cities are shown in some cases to be efficient producers of economic[3], social[5], and environmental[6,7] benefits. So, it is no surprise, that some of the solutions developed by private and public sectors are to build microcosms of cities in urban areas known as *mixed-use developments*[8,9].

Many definitions exist for mixed-use development, but it is commonly defined as a well-planned development that integrates more than one function—or specifically, more than one program as it is referred to by architects and urban planners (e.g., commercial and residential)[10]. They have been a fixture of both urban and sub-urban areas to address density, optimize land use, and are seen as an emerging trend in future cities[11]. We hereby refer to mixed-use developments that consist of commercial, residential, and community (public) facilities closely integrated with transportation networks as an integrated development. However, it is worth noting that the term 'integrated development' is not well defined. Our study uses the term 'integrated' to define both outward-integration (to the larger spaces such as a district) and inward-integration (e.g., multi-level and street-level connections among all spatial and functional parts of the development). Integrated developments show promise to be part of the solution to address the UN SDG goals. They have the potential to improve social interactions, reduce demand for transport and other infrastructures, and improve the safety and vitality of the community[12]. However, research on or related to mixed-use developments is sparse[10,12,13].

Beyond better land-use planning for the built environment, city planners are adopting data-driven approaches for the design of the so-called 'smart city'[14,15]. The smart city concept can tap the potential of distributed multi-sensory networks—including solutions based on the Internet of Things (IoT)—to measure various features of a city such as traffic, energy, waste, environment, and people[16]. These measurements constitute 'big data' for the city, which can be assimilated to improve the real-time efficiency of a range of operations and to build better cities for the future. A similar approach has been developed specifically at the building level—the so-called 'smart building' framework[17,18].

Hence, data collected from human behavior in a given city or building is related to spatial and environmental measurements that can be harnessed to improve a range of features of the built environment such as walkability, ease of navigation, and convenience[19]. This has been traditionally performed for street networks from the city scale down to the neighborhood scale. It is also not uncommon to find studies that extend those concepts down to the building scale[20–22] or even the floor level[23]. These studies generally represent the structure of the built environment as a network and the human behavior related to it as a dynamic process taking place in it. In Architecture, this network representation is often considered within the space syntax framework[24,25], which has frequently been cited as being limited in scope due to its dual graph representation[26,27]. Urban researchers and geographers have adopted a more general network-theoretic approach to studying the urban system and its components using *network science*[28]. Over the last two decades, network science has grown explosively and imposed itself as a powerful abstraction that enables the representation and study of many types of systems as networks, and that across a wide range of disciplines (e.g., human, ecological, and robots)[29]. For details on the difference in approach between space syntax and other network representations of the built environment, we refer the reader to the work by Porta et al.[27].

Currently, there is a lack of a general framework to analyze the effectiveness of integrated developments. Unlike a sprawling city or a mixed-use street, the integrated development introduces heterogeneity in space and function in a high-density built environment that is clearly more complex than single-use developments[30]. Another key challenge is the lack of appropriate data that can help test or evaluate a given design hypothesis for an integrated development. Evidence-based or data-driven design[31] is a growing trend in urban development based on the premise that we cannot improve what we cannot measure. The same can be said at the integrated development level, hence pointing to the pressing need to gather data, process and analyze it.

Human movement in a building is commonly referred to as circulation. The study of circulation is an essential part of the design process in Architecture that aims to improve legibility, navigation, accessibility, safety, and optimal connectivity[20]. The relationship between human movement and the built environment (city[32], street[33], building[20,34], or a floor[23]) has been extensively studied using space syntax and other network-theoretic methods[35–37]. This is done so using socio-spatial theories or based on movement data. However, movement data in integrated developments, or more broadly in mixed-use developments, are sparse. Chang[30] collected movement data (adjacent and path) from 300 users in an integrated development (Barbican, London, U.K.) and a sprawling mixed-use development (South Bank, London, U.K.) based on visual observations to study individual route choice and movement behavior. In addition, Willis et al.[33] collected movement data from 2613 participants in three mixed-use streets in the U.K. based on video-recorded observation to study pedestrian movement. These studies[30,33] belong to the rare cases of works that analyzed movement data in mixed-use developments. It is apparent that the scant presence of adequate human movement data[38] hinders our ability to test design hypotheses in mixed-used developments. Our study introduces movement data that tracked 51 participants over a long period of time (1–5 days) in a free-living environment with high temporal ($\sim$ 1 Hz) and spatial resolution ($\sim$ 20 m) within a new state-of-the-art integrated development in Singapore. Human Activity Recognition (HAR), such as tracking and identifying human movements precisely in a three-dimensional (3D) environment, requires a suite of sensors (e.g., barometer and IMU[39]) and involves complex data processing[40]. This study uses strategically placed Bluetooth beacons (environment) and mobile applications (peer) to finely track user movements.

We adopt a general network-theoretic approach similar to urban researchers and geographers[41]. We study network representations of factors in the built environment that are not embedded in space, i.e., besides choosing built environment fixtures as nodes and their adjacent pathways as edges, we represent the spatial functions and their relationships as a weighted network that is not necessarily embedded in space, and we also aggregate spatial components for abstract analysis[42]. This method is similar in essence to the approaches used to study other networks like social or ecological networks, where more abstraction of network components results in a

better understanding of the network process and its system-level operation. This framework and methodology are especially relevant in studying integrated developments that introduce complexity in the relationship between humans, space, and function. As mentioned previously, this method is different from space syntax methods traditionally used in architecture. We refer the interested readers to work done by Zhang and Chiaradia[43], for the application of space syntax methodology to understand three-dimensional mixed-use areas using user movement.

Hence, analyzing the factors that influence user movement in an integrated development pave the way towards a better understanding of the intricate relationship between human activity and the built environment. To improve the design of these developments, data-driven approaches have the potential to uncover all the factors affecting human movements such as the human aspect, the spatial aspect, location, time, and climate[19,44]. On the one hand, spatial factors include geometry, connectivity, visibility, spatial integration[34], centrality[45], intelligibility[20], legibility, walkability[46], and spatial function[23]. On the other hand, human factors include familiarity with the environment[30], spatial cognition, user profile, and user preferences[33,47]. In this study, we focus on two spatial factors, namely topology and function. The term 'function' refers to space utilization i.e., primary use of space such as commercial, community, corridor, or vertical facilities.

In both space syntax studies and the general network-theoretic approach, the appropriate use of distance measures is an ongoing topic of research. The distance between two nodes in a network can be given by its topological (number of hops), Euclidean (straight line metric distance or distance along the network), angular (sum of deflection angles), or hybrid distance (combination of the above). Analyzing the road network of the UK using vehicular data, Serra and Hillier[48] showed that angular distance is a better predictor of vehicular movement than Euclidean distance. However, the study also showed a very high correlation between the betweenness centrality measures when using angular and Euclidean distance at small spatial scales ($\sim$ 2 km). Similarly, Zhang et al.[49] analyzed the road network of Shanghai, China, and found that angular and Euclidean distances are highly correlated (80–90%). Moreover, Zhang and Chiaradia[50] studied the three-dimensional pedestrian network of Central, Hong Kong using topological, Euclidean, angular, and hybrid (Euclidean and Angular) distance, and showed that the betweenness centrality correlation (spatial measure and movement) is the same for all except the topological distance. The study however also shows that this is not the case for closeness centrality correlations. More research is required to understand the effect of studying the relationship between spatial measures and user movement using different distance measures. In addition, Cooper[51] notes that users unfamiliar with a route tend to take the path with few angular turns. On the other hand, familiar users are prone to select routes that minimize the Euclidean distance. As the participants tracked in our study are residents, employees, and frequent users of the space, they are very effectively familiar with the building and its configuration. Therefore, the Euclidean distance has been selected as the natural distance metric.

It is worth noting that the Euclidean distance along the network is different from the classical Euclidean distance, where the latter is calculated by drawing a straight line between one node to another while ignoring the presence or absence of a direct physical connection between them. Our study thus uses the Euclidean distance along the network (a.k.a. routing distance) to reflect the realistic distance a user will traverse and in some cases, the network is purely analyzed by its topology. In addition, the choice of Euclidean distance along the network is appropriate for familiar users known to minimize the distance

We ask ourselves the following question: how do topology and function influence user movement/circulation in a complex integrated development? User movement data is collected in Kampung Admiralty (KA), Singapore's first integrated public development that brings together a mix of public facilities (healthcare, gardens, and community spaces), commercial facilities (shops, banks, and restaurants), and a residential complex[52–54]. KA serves as our case study to demonstrate the promise of a data-driven approach in this framework and the power of a network representation to understand human-built environment relationships in a high-density complex environment. Our contribution to integrated development analysis is limited to KA's inward-integration property while its outward-integration is only studied up to the multiple entrances that seamlessly integrate into the rest of the district. The space beyond KA's entrances is out of the scope of this study. Our results can be summarized as follows: (1) we establish a clear quantitative relationship between building topology and user movement, (2) we uncover the details of the interplay between topology and function in influencing user movement, and (3) we quantify the magnitude of movement flow within and between the spatial and functional components of the integrated development.

Results

The study aims to understand the impact of building topology and function on the user movement within. While the KA spatial network is an abstraction of the building topology, the user movement can be studied as flows across this connected network. To uncover hidden patterns of user movement, we aggregate the movement data according to specific floor groups and spatial functions (refer "Methods" section).

The first section quantifies the relationship between network topology and user movement. The second section looks at the interplay between connectivity and function through user movement data and the underlying spatial network. The third section analyzes the homophilic and heterophilic movement flow across the vertical components of the building and spatial functions and explores the impact of their relationship on user movement. In this last section, we also use the origin-destination matrix to understand the influence of topology and function on the beginning and end of each trip taken by the user and finally calculate the node entropy to look at the diversity of the movement flow across the spatial and functional groups (refer "Methods" section).

Influence of spatial network topology on user movement

As a first step, we attempt to define a clear quantitative relationship between the topology and user movement.

Various centrality measures can determine a particular node's importance in a network. The description of the selected node centrality measures and their relevance to the built environment are explicitly listed in Table 1. We calculated the Spearman's rank correlation coefficient between node centralities (in-degree, weighted closeness, and weighted betweenness) and movement flow to understand the relationship between the topological network properties and the network process—i.e., the user movement flow. A moderate correlation ($\sim$ 0.4) is observed between all of the topological properties and the movement flow with statistical significance (p-value $\leq$ 0.001) (Table 2). This shows that the network topology itself has a moderate influence on how people move about throughout KA. Previously, these results have been observed for street networks and other transportation networks[55–58]. At the building level, these results are consistent with other works where building configuration is quantitatively shown to influence user movement[34]. However, an explicit quantification of the magnitude of this topological influence, as shown here at this scale, is still rare[59]. As we break down the KA spatial network and examine its sub-graphs, the correlation values change for different floor groups. Level-1 and L-2 have a high correlation (0.5–0.8) between the given topological properties and movement flows, while car-parks show low-to-zero correlation (Table 2).

The topology's moderate influence means that many other parameters drive the movement flows, including function, user preferences, and other properties that are not studied in this work such as weather conditions, spatial enclosure (partly indoor/outdoor), spatial cognition[60], and facilities (e.g., presence of sitting places)[59].

Interplay between connectivity and function

In the previous section, we established a quantitative relationship between topology and user movement. In this section, we analyze and demonstrate the influence of topology and function broadly. When a user moves from one location to another in an integrated development, she moves across various spaces with particular functions. The user is also moving along the designed topology of the built environment across the horizontal and vertical directions. Does a user move from one space to another due to spatial function, or does better connectivity allow for it? The complex interplay between these factors influencing the movement volume is illustrated and analyzed. This section takes both a qualitative (based on on-site observation and knowledge of the building) and a quantitative (aggregated movement data and weighted spatial network visualization) look at the results.

The number of movement inflow for a given node in the KA spatial network is studied according to floor group and further sub-divided by function type as shown in Fig. 1. The aggregated distribution shows rich movement flow across the vertical and horizontal dimensions. Level-1 (ground floor) has the highest inflow due to its accessibility from the ground (entrance), large open plaza (social space), and its contribution as the main commercial street with shops and restaurants (commercial). This particular diversity of function is clearly reflected in the significant movement flows in the above-said function categories. Level-6 connects the residential and commercial components of KA and sees the second most movement inflow, similar in magnitude to the entire residential building that has seven floors. Residential movement is overwhelmingly in the vertical dimension as there is little horizontal spread on residential floors. This is followed by Level-2 that only consists

Node centrality	Description	Relevance to built environment
In-degree	Number of incoming connections to the node	A node with high in-degree centrality has many incoming connections and hence it is said to have high connectivity locally
Closeness	Average distance of a node from all other nodes	A node with high closeness centrality is said to be the optimal node (to minimize distance) to start to reach all other nodes. For example, designers can place an entry point or key zone with high closeness centrality for optimal reachability to/from all other nodes
Betweenness	The extent to which a node lies on the shortest path between other node pairs	A node with high betweenness centrality is said to be central to all the traffic flowing between other node pairs. For example, if a designer wants a key zone to be placed in such a way that most people walk through it when traveling between any two locations, the node can be designed to have high betweenness centrality

Table 1. Node centrality measures and their relevance to studying the built environment.

Floor-groups	In-degree	Closeness (weighted)	Betweenness (weighted)
All nodes	0.402***	0.387***	0.391***
Car-parks	0.094	0.055	0.14
Level-1	0.571**	0.337	0.655***
Level-2	0.541*	0.782***	0.595*
Level-6	0.37	0.203	0.216
Rooftop garden	− 0.023	0.486	0.607*
Residential	0.438*	0.334	0.307

Table 2. Spearman's rank correlation coefficient results between topological network centralities and mobility (incoming flow) for all and each floor-group. The closeness and betweenness centralities are weighted by routing distance. ***p-value $\leq$ 0.001; **p-value $\leq$ 0.01 ; *p-value $\leq$ 0.05.

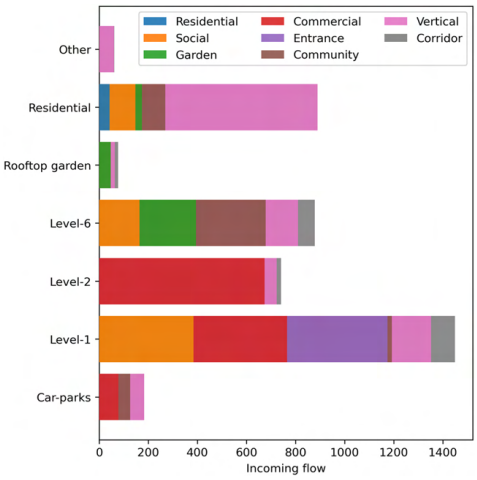


Figure 1. Distribution of inflow by floor-group and function categories.

of food and beverage stalls with accompanying sitting areas, which nevertheless see a significant movement flow. The car-parks (Basements 1 and 2) and rooftop gardens (Level 8 and 9) form the two corners of KA's vertical structure, and unsurprisingly they exhibit little movement flow. The car-parks consist of commercial spaces such as supermarkets and banks. The majority of the user movements within those floors is associated with these commercial spaces.

The usage of garden spaces on Level-6 can be unavoidable if a user needs to cross between the residential and the commercial tower on that floor. This is, however, not the situation for the rooftop garden where the user movement indicates a deliberate use of garden space. In fact, the users have vertically traveled to the top floor of the building with the sole intention to do so. The spatial network of KA is shown in Fig. 2 where the nodes

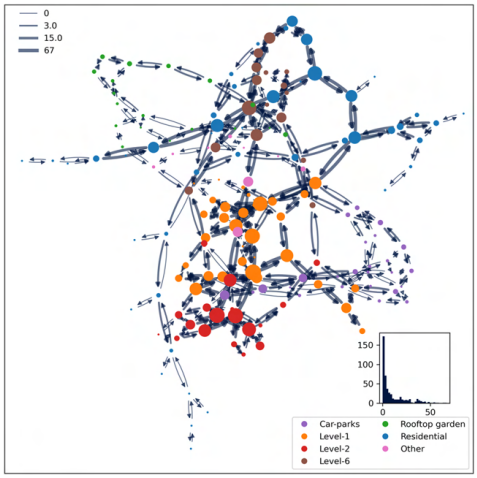


Figure 2. KA spatial network. Nodes are colored by floor group and sized by movement inflow. Directed edges are sized by the number of movement flows in respective directions. Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50th percentile, 75th percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

are colored by floor group and sized by inflow. The spatial network tapers at the two spatial ends of the KA building. Both connectivity and function drive the movement distribution between those two ends (car-parks and rooftop gardens). However, the Level-6 connection between the two mid-rise towers has allowed for higher movement flow, where connectivity has influenced better movement flow than otherwise possible with floor function alone. This is seen by the slight movement flow on Level-7, which has a similar function space grouped under the category 'Other' in Fig. 1.

The movement patterns in the six-floor groups are analyzed and presented in this section (Fig. 3). The vertical streets (lifts, escalators, and stairs) form the KA spatial network sub-graphs cornerstones. They serve as the main access points for elevated or basement floors. The movement distribution is concentrated at some point of the network sub-graph for car-parks, Level-2, Level-6, and residential buildings. However, this concentration is not always a property of the network itself, i.e., only some nodes of these sub-graphs have mobility flow proportional to the node degree, such as for instance the center point of Level-1 and Level-2. In some cases, the movements are clearly influenced by function, e.g., commercial nodes of car-parks and the commercial side of Level-6. Nevertheless, there is significant movement flow at multiple points of the sub-graphs showing good circulation throughout KA in most cases.

In summary, this section looked at the aggregated distribution of movement data and weighted spatial network visualization to illustrate the influence of connectivity and function on user movement. Qualitatively, it can be inferred from the data analysis that both factors show varying degrees of influence.

Homophilic and heterophilic movement flow

So far, we have shown evidence of the impact of topology and function on user movement. In this section, we look at this influence in more detail and in a more quantitative way. How intense is the movement flow within and between the spatial and functional components of the integrated development? We specifically aim to study the influence of the relationship between the defined spatial and functional groups on user movement. This is done by studying the aggregated movement data based on two types of movement data—adjacent movement (studied so far) and origin-destination of a trip.

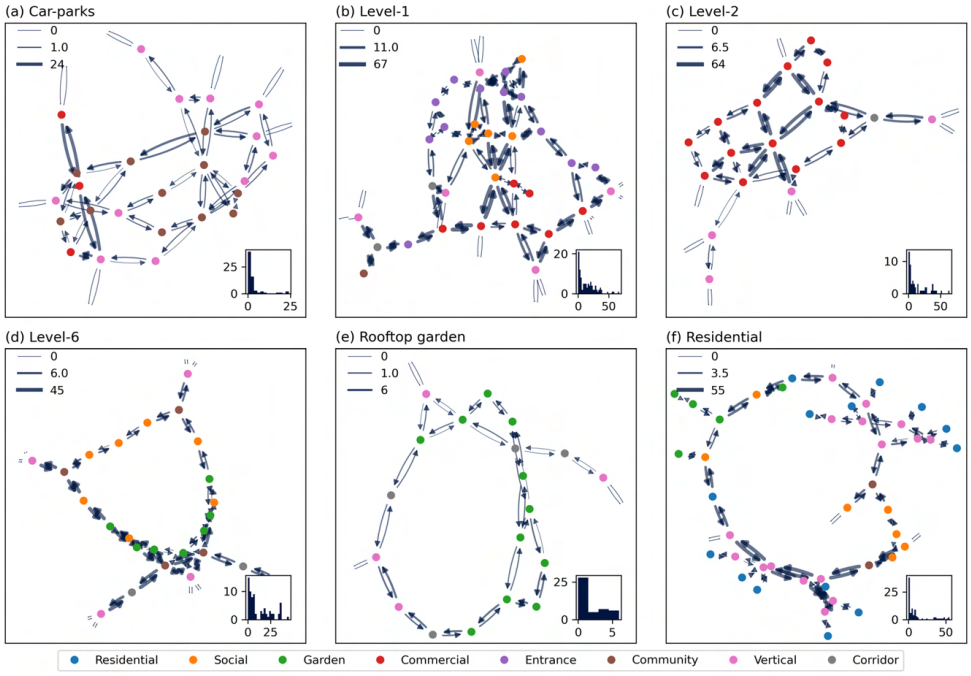


Figure 3. Subgraphs of the spatial network. (**a**) Car-parks (B1 and B2), (**b**) L1, (**c**) L2 (food court), (**d**) L6, (**e**) rooftop garden (L8 and L9), and (**f**) the two residential towers. The colors indicate function categories. Bottom right plot: Histogram of flow count. Top left: Legend indicating flow size for minimum, 50th percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

Adjacent movement

The spatial network is aggregated by floor-group and functional-group. Table 3 shows the proportion of movement within (internal) and between (external) the six floor-groups and eight function categories. The internal flows (homophilic) account for 80% to 90% of the nodes' movement in the floor-group; i.e., only 10% to 20% of the flows left each floor-group. Many of the function groups have the majority of the movement flowing externally. This may be because they are also physically scattered in space. An exception to this is the vertical and commercial function spaces. These groups are physically clustered and interconnected, thus allowing a higher proportion of the movement flow to remain internal.

Figure 4 presents the external movement flow of people by means of the widths of edges, and the size of internal flows is presented as the size of nodes. The external flow across the floor groups is more or less equal in volume between L-1, L-2, L-6, Residential, and the car-parks (Fig. 4a). However, the internal flow is highest on L-1, followed by the residential region and L-2. The lowest internal user movements are seen, not surprisingly, in the car-parks and rooftop garden. This is also clearly seen in Table 3. The external flows between different functional groups are more or less equal. An exception is a low movement between the community and commercial spaces. This only indicates that those two different functions are mostly physically separated, and hence only a few adjacent movements were possible. The residential units and corridor have zero internal movements because they are physically separated, and their sub-components are only connected through vertical function spaces like lift lobbies. The vertical spaces are the central function space connected to all other spatial functions, thus driving the majority of the external movements. Loosely speaking, these lift lobbies operate like hubs in KA. It is worth noting that the mobility data is not studied by hour of the day, weekday, or weekend due to the small

Grouping	Category	Total	Internal	External
Floor-group	Car-parks	236	0.78	0.22
Floor-group	Level-1	1638	0.89	0.11
Floor-group	Level-2	877	0.84	0.16
Floor-group	Level-6	693	0.86	0.14
Floor-group	Rooftop garden	85	0.91	0.09
Floor-group	Residential	986	0.90	0.10
Function category	Entrance	402	0.50	0.50
Function category	Social	586	0.33	0.67
Function category	Commercial	1277	0.80	0.20
Function category	Community	351	0.18	0.82
Function category	Garden	306	0.34	0.66
Function category	Residential	42	0.00	1.00
Function category	Vertical	1496	0.72	0.28
Function category	Corridor	195	0.01	0.99

Table 3. The proportion of internal flows and external flows for floor-groups and function categories.. 'Residential' in the Floor-group refers to both residential towers, while 'Residential' in the function categories denotes the individual apartments of the two residential towers that exclude vertical transportation and corridors.

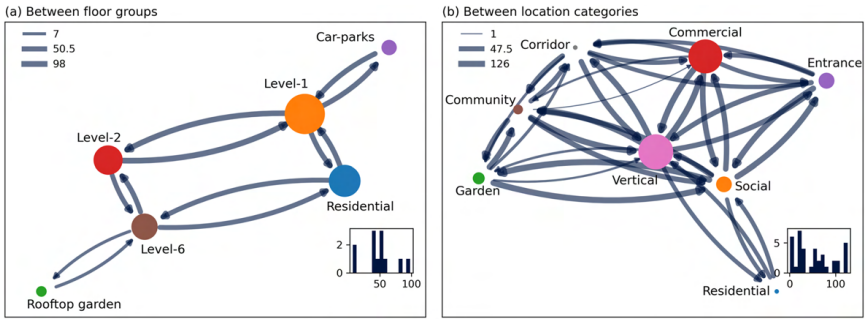


Figure 4. Networks of aggregated flow. (**a**) By floor-groups and (**b**) by function categories. The node sizes indicate internal flows (self-loop). Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50th percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

sampling size in such temporal divisions. Such intricate analysis, however, could enrich this study by showing temporal patterns of human movement and its relationship to the built environment (see "Discussion" section for more details on the limitations of our study, including the small size of the representative population).

The vertical streets are one of the main pathways in KA integrated development. However, the high internal flow in commercial areas also indicates that the movement is heavily 3-dimensional. These results on user movement flow highlight the intricate relationship between functional and spatial units of the groups, and reveals the key players driving those movements.

Origin and destination of a trip

The previous section dealt with analysis of the homophilic and heterophilic movement flows. These results are limited since they are based on using adjacent movement data. Specifically, they ignore the two most important parts of user movement, namely the origin and destination of each trip. By studying the origin-destination matrix and its network representation, we aim to deepen our understanding of the relationship that exists across the vertical components of the building and spatial functions. The start and end of a trip bring the components of these groups closer in context as we see the user movement in full circle, and thereby sheds light on the destination preference of a user.

Figure 5 shows aggregated network graph with each directed edge indicating an origin and destination of a trip and its size denoting the volume of trips between them. The size of the nodes indicates the number of trips whose origin and destination are within the same node. Figure 6 shows the Origin–Destination (O–D) matrix with the diagonals denoting the internal trips. Most trips have originated or ended at level-1 (see Fig. 5a). The

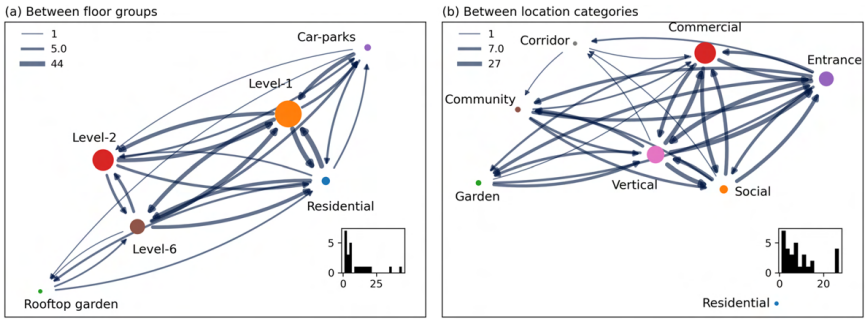


Figure 5. Networks of aggregated flow based on origin and destination of a path. (**a**) By floor-groups and (**b**) by function categories. The node sizes indicate internal flows (self-loop). Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50th percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

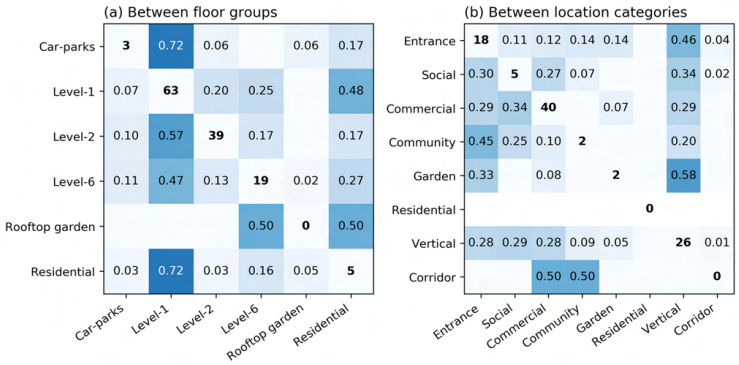


Figure 6. Row normalized origin-destination matrix. (**a**) Aggregated floor-groups network, and (**b**) aggregated function categories network. Reading the OD matrix: The diagonal values indicate the number of internal flows within each group. The row normalized values indicate the proportion of flows from one origin to every destination. For example, 0.72 in the first row indicates that 72% of movement originating in car parks ended at Level 1..

largest destination of trips originating from car parks is level-1, but the users have also reached level-2, residents, and rooftop garden (see Fig. 6a). The trips starting from the residential building have also majorly ended at level-1 and have reached every other floor. The large volume of users departing to and from the rooftop garden is from the residential building. As already mentioned, the residential units and corridor have zero internal movements because they are physically separated, and their sub-components are only connected through vertical function spaces like lift lobbies. This is clearly visible in Figs. 5b and 6b. The Level-1's connectivity to the rest of the neighborhood and various function spaces allows most of the trips to start and end in level-1 as shown in Fig. 5a. Nevertheless, the edges between the vertically discrete spaces show that KA integrated development encourages a large volume of trips (and thus destination) within.

Nearly 50% of the trips starting at the entrances reach the vertical spaces, thus showing that those trips are eventually destined for elevated spaces (see Fig. 6b). The trips starting at entrances and vertical spaces see the most diverse destinations. Trips with origin and destination within the same function spaces are seen primarily in commercial and vertical spaces. This is similar to the results seen in the previous section using adjacent movement data. Similar to the floor group, the aggregated function group network is highly connected, thereby indicating trips with diverse origin and destination within KA.

Vertical nodes are not typically considered as an origin or destination but as transition nodes. However, they are present here as origin and destination due to the fact that—(1) vertical facilities are numerous and placed very close to several entrances of KA, and (2) exiting vertical nodes on higher floors can lead to private locations like residences that are not monitored. Considering vertical nodes as part of the O–D matrix enriches our three-dimensional study of mobility and provides an opportunity to understand the role of vertical nodes in facilitating trips within KA. Please refer to Supplementary Fig. S1 which features the O–D matrix with the vertical node removed.

Diversity of movement flow

Table 4 shows the normalized outgoing and incoming entropy, which respectively represent the diversity of outflows and inflows. A high value of outgoing entropy ($H^{out} \simeq 1$) indicates evenly distributed outflows to all other nodes, whereas a low outgoing entropy ($H^{out} \simeq 0$) indicates the outflows were all going to the same destination. The highest flow diversity is seen in the outflow of Level-6 and inflow of car-parks and Level-1. Most other floors have low to moderate diversity of trip origin and destination. The car-parks have a moderate outgoing entropy and high incoming entropy, suggesting that the incoming trips have started from diverse locations while the diversity in the destination is low. The rooftop garden has the lowest outgoing entropy, thus indicating trips with few destination choices. The highest trip diversity is seen at the entrance and vertical spaces, as previously shown in the O–D matrix (see Fig. 6).

The homophilic and heterophilic analyzes of movement flow show a strong movement within floors and function groups. By looking at the weighted spatial network at an aggregated level and considering adjacent movement and trip data, we quantified the influence of topology and function on the user movement.

Discussion

This study is aimed at identifying and understanding the factors that influence user movement in an integrated development like KA, where residential, commercial, and public facilities are distributed in a vertical, high-density built environment. We specifically focused on two factors—topology (built environment layout and spatial network structure) and function. With Kampung Admiralty as a case study, we show that network-theoretic methods offer a powerful way of looking at data-driven designs at the building scale by associating the network process (user movement) with the network structure (built environment layout). We asked and answered specifically the following three research questions: (1) "What is the quantitative relationship between topology and user movement?", (2) "Does a user move from one space to another due to spatial function, or does better connectivity allow it?", and (3) "How intense is the movement flow within and between the spatial and functional components of the integrated development?".

First, our results reveal a $\sim$ 40% correlation between the centrality measures of the KA spatial network and user movement, thereby indicating that the distribution of flows followed the topological network structure to a moderate level. Previous studies at the country- and city-scale suggested that the correlation between the daily

Floor-group	H^{out}	H^{in}	Function category	H^{out}	H^{in}
Car-parks	0.477	0.738	Entrance	0.732	0.691
Level-1	0.671	0.709	Social	0.649	0562
Level-2	0.641	0.547	Commercial	0.614	0.600
Level-6	0.729	0.629	Community	0.605	0.561
Rooftop garden	0.387	0.530	Garden	0.427	0.492
Residential	0.505	0.615	Residential	0	0
			Vertical	0.716	0.690
			Corridor	0.333	0.500

Table 4. Normalized outgoing entropy (H^{out}) and incoming entropy (H^{in}) of the floor-group and function categories based on origin and destination of a path.

human movement patterns and network metrics was higher at about 60%[55,56]. Urban planners and designers can benefit from an accurate and quantitative understanding of what drives people to move and spend time in locations. These are already widely studied in human mobility studies at the city scale for optimizing the allocation of facilities and efficient design of the built environment[61]. The question of which factors contribute to what proportion of movement can precisely guide the designers to focus their effort on factors that are proven to work. More research and data-driven post-occupancy studies are required to study the precise influence of all factors that affect user movement at different spatial scales, to understand both universal and subjective behaviors (e.g., design-dependent, and population-dependent).

Second, the analysis of the KA spatial network and its sub-graphs show a varying degree of intertwining influence by topology and function on user movement. Our results indicate that a user is driven to specific areas due to spatial function. However, the route choice of the user is influenced by connectivity. Traditionally in planar studies, route choices can be influenced by the shortest or the fastest route. In our case study, the user preferences for the route are often influenced by spatial functions along the route like garden spaces and plaza areas. Connectivity between the different nodes can thus be leveraged to create the desired quality of the space, including allocation of desired spatial function, and in the process influence the user route choices.

Third, by exploring the movement flow between and within various functional and spatial components of KA, we quantify and reveal a robust 3D circulation. High building circulation is an essential design intent of mixed-use developments, especially in developments like KA, where most functions are distributed across horizontal and vertical dimensions[62]. Our results also show that the spatial function in KA is sufficiently heterogeneous to allow for a large volume of trips with diverse origins and destinations. Vertical streets (stairs, escalators, and elevators) are built to be the backbone of high-rise developments. Hence, it is no surprise that we found these vertical streets to be the hub of the KA spatial network, forming the main pathway of the integrated development. Although observations and theories from previous studies have hinted at some of the inferences drawn here[45], our study offers a quantitative assessment of the complex relationships between the functional and spatial components of an integrated development through the perspective of human movement flow based on high spatiotemporal data.

One limitation of our work is the relatively small number of participants, due to the constraints of the peer-to-environment sensing that requires the active recruitment of volunteers willing to take part in a long-term study. More passive systems like Wi-Fi sniffers could help track a larger population but inevitably lead to serious privacy concerns. Moreover, the study's data collection period happened in February 2021, which is a recovery period from one lockdown due to COVID-19 in Singapore. Although most of the daily activities were starting to be resumed during that time, the Singapore Government advised the public to reduce social activities and gatherings. Thus, the human movement patterns in this study could also have been affected by the preventive measures in place due to the COVID-19 pandemic. We have not identified and quantified all the factors that influence user movements, such as the human aspect, spatial aspect, location, time, and weather[19,44]. Hence, more data and analysis are required to fully understand user movement in complex built environments. This also includes constructing co-presence networks of users to study their social interaction and recording user occupancy data to study space utilization. The use of appropriate distance measures (topology, Euclidean, angular, and hybrid) to study spatial networks[48,50] is an ongoing research. Our study is potentially limited by the use of Euclidean distance and more research is required to understand the impact of using particular distance measures on the phenomenon studied.

Our study does not incorporate design features that can differentiate design classes between developments that can potentially have the same geographic or abstract network representation. This is especially useful when a comparison between different development's spatial design and user movement is explored. The network-theoretic approach offers the potential to include distinct design features as node or edge properties to make such comparisons. However, more data-driven research is required to test the effectiveness of this approach. In addition to exploring design-dependent properties, population-dependent features can be incorporated into the social network of users that traverse these spaces. This social network can be encoded with their co-presence[63] (or contact patterns—time and space in which two users are present simultaneously), psychological profile, spatial cognitive profile, and user preferences[64]. The simultaneous study of the interaction between the spatial network of the built environment and the social network of users (known as a socio-spatial study) is a hot topic of research in human mobility studies[65].

Future cities may need to grow dense and vertical to meet the UN-SDG goals for 2030. Urban planning and design for these cities is tied to our understanding of how people use the built environment. Integrated development is not an independent urban element; on the contrary, it is fully integrated into districts and neighborhoods. This anchoring in a given district/neighborhood has to be taken into account (e.g., mass rapid transit, linkways, bridges). The lack of understanding of human movement in the built environment such as integrated developments can lead to poor designs of related buildings and ineffective integration within the overall urban space. More research is required to expand this study beyond the building to the district scale. Liang and Kang[66] list four critical challenges for the urban planners and designers to use the insights gained from spatial network analysis: (1) data openness and privacy, (2) lack of direct policy implications, (3) lack of civic, communicative, and collaborative engagement, and finally (4) difficulty to automate visualizations and integrate with Geographic Information System (GIS). Research and practice related to these four challenges are key to the development and use of data-driven planning and design approaches from the building all the way up to the city scale. Only then, future cities can grow vertically and accommodate higher densities while remaining livable.

Methods

This study uses a data set obtained from a human-tracking experiment carried out at KA for over one month (February 2021) involving 51 distinct participants. This experiment has been designed to collect movement data from regular users of KA as they move around in a free-living environment. The following sections describe the data collection system and processing, movement data, study site, and the derived spatial network.

Data collection and processing

The data collection system is based on a peer-to-environment Bluetooth localization method that consists of three components: (1) stationary low-energy Bluetooth beacons (Kontakt Smart Beacon Pro BP16-3), (2) custom-built data collection mobile app (iOS and Android), and (3) an access to the AWS cloud server. One hundred twenty-four beacons are placed in locations of interest within and around the KA building perimeter. They transmit their unique ID with a sampling interval of 10 Hz, which is scanned and stored by the participant's smartphone, and the information is eventually uploaded to the cloud server.

The beacon placement strategy is based on the varying sizes (and signal blocking physical features) of the location grid, which called for the customization of the beacon transmission range (10–50 m). That also allowed us to reduce the signal interference between the beacons and limit the number of beacons detected at each location grid to 1–3 units by design. In turn, this enabled us to implement a straightforward and accurate localization method within the building. This localization is based on a simple transformation table that relates the scanned beacon with the highest Receiver Signal Strength Indicator (RSSI) to a single location grid/zone. To create this table, 4 h of training data are collected using two smartphones (Google Pixel 4a and Samsung Note 8). This ground truth data has been acquired by two researchers carrying these mobile devices while covering systematically the process area of KA, the data being recorded with a sampling rate of 1–0.2 Hz.

The exact same path was recorded twice with the two smartphones. Subsequently, a Naive-Bayes machine learning model is trained on 60% of the training data collected from one mobile and tested on the remaining 40% collected from the second mobile phone. The model achieved a classification performance of 100% accuracy at floor level and 80% accuracy at the node level. Furthermore, it reached a 93% accuracy at a slightly coarser location prediction involving the nearest neighbor. The nearest-neighbor level prediction accuracy of 93% is suitable for our study as our analysis consists of grouping the individual location zones and studying the movement data by aggregation. The transformation table that maps the beacons (with the highest RSSI) to each location grid/zone is thus generated and applied to the mobility data set.

Description of movement data

The data set consists of movement data collected from regular users of the Kampung Admiralty building with duration of recording per participant varying from minimum one day up to a month in February 2021. Volunteers were recruited by representatives in booths or through posters placed around KA. They were offered vouchers worth S$30 in two installments—during recruitment and later at the end of the data collection effort—as an incentive to keep the mobile app installed. Each participant is tracked continuously for the entire duration of the experiment (i.e., day and night). However, their location can only be determined if they are ambulating within KA. To reduce the impact of a few users with long tracking data from dominating the movement data, the maximum number of days per user is restricted to 5 days (refer Fig. 7a for a detailed breakdown of the number of days per participant. The group of 51 participants comprises 21 males and 30 females between the age of 17 up to 90 years old, with 50% below 59 years old (refer Fig. 7b for a detailed breakdown of participant age). The high elderly demographic is representative of KA's regular users whose residential complex is meant to house seniors and includes community facilities like Active Ageing Hub. Of the 51 participants, 18 are residents, 9 are employed within KA, and 24 are frequent visitors. A total of 131 participating days (i.e., sum of 1-day worth of data for all users) from 51 participants were extracted from the KA data set.

The trip data is derived from the time series of predicted locations using the transformation table. Each trip data is a path through the KA spatial network. This also allows us to fill in missing nodes (location) in a given

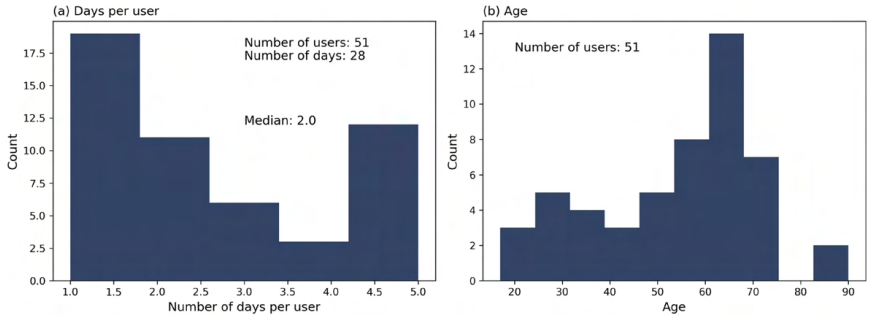


Figure 7. KA Movement data—Histogram of (**a**) number of days per user and (**b**) age.

path by reconstructing it using the shortest path between the known locations for a conservative estimate. The predicted series of nodes in a path has a length of 3 nodes or less for 87% of the missing patches applied. Hence, even though this trip prediction method partly depends on the topology (shortest path calculation), the above numbers suggest that it does not significantly impact our study on the correlation between topology and user movement. A total of 581 paths were derived, of which 361 are continuous in time with a pause time of fewer than 5 min at each stop (i.e., $\leq$ 5 min). These 361 paths are defined as a trip in this study, whose origin and destination form the entries of the Origin-Destination matrix. The derived 581 paths encompass 510 paths from weekdays and 71 paths from weekends. A brief comparison between weekdays and weekends is shown in Supplementary Figs. S8 and S9, and Supplementary Tables S5 and S6. It is important to note that the data obtained on weekends is too small to make a meaningful comparison.

A pause of 5 min is selected based on the following consideration: (a) a pause should reflect a destination with sufficient time to perform activities in KA such as visiting a store, or making inquiries in community facilities, (b) allow minimal time for users to move from one location to another, (c) account for wait times in lift lobbies or potentially missed localization. In human mobility research, a pause time of 10–20 minutes is considered typical to define a trip for traveling across cities or large geographic areas[67]. However, KA's compact space and multiple opportunities for destinations require a relatively smaller pause time. Please refer to Supplementary Figs. S2–S7 and Supplementary Tables S1–S4 for results obtained with a pause time of 10 min. More information on the study site and the KA spatial network is given in the following sections.

Study site

Kampung Admiralty is Singapore's first vertically integrated public development (11 stories and 2 levels of basement, completed in May 2017) that brings together a mix of public facilities and services under one roof (see Fig. 8a)[68]. As the increasing urban density in Singapore demands creative ways of intensifying land-use effectively in the vertical dimension, the elevated and layered urban design and architecture approach to the project led to transforming the 0.9 ha site into a dynamic vertically integrated mini-neighborhood for the community. KA's designers used a 'sandwich layered' approach to vertically join various urban functions into that 0.9 ha area. This enhanced the mixed land-use management and living experience of the residents of KA and its vicinity. As an integrated development, the KA is immediately adjacent to a bus stop and a train station (mass rapid transit: MRT station Admiralty). We consider KA to be a high-density development due to the following reasons—(1) population density: KA is placed in a residential zone, the Woodlands East sub-zone, where the population density is very high ($\sim$ 38,000/km^2), (2) Floor Area Ratio (FAR): KA's FAR is 3.6, a relatively high ratio for a residential zone, and (3) integration of multiple functions: KA integrates various amenities and services within a relatively compact area, a characteristic of high-density developments, aiming to optimize land use and promote a sense of community.

KA consists of two main parts: the community side and the residential side. The latter comprises two residential towers (the plus/cross shape towers in Fig. 8). The rest of the area forms the community side, which includes multiple types of facilities such as childcare, elderly care, car-park, hospital, shops, bank, and restaurants (refer to Supplementary Fig. S11 for the distribution of these functions across the floors).

KA spatial network

KA provides a mix of public and common spaces distributed across multiple levels and a mix of spaces with recreational facilities. The whole public space of KA is subdivided into a total of 165 space units (Fig. 8b), and the adjacency connections are linked as shown by the edges (Fig. 8c). A *node* in the KA network is defined as a programmed space with defined boundaries. An *edge* of the network is directed, and it is formed between two reachable adjacent spaces. Thus, the KA spatial network has 165 nodes and 476 edges (Table 5). The routing distance between the centroid of any two nodes is assigned as an edge weight. The KA spatial network consists

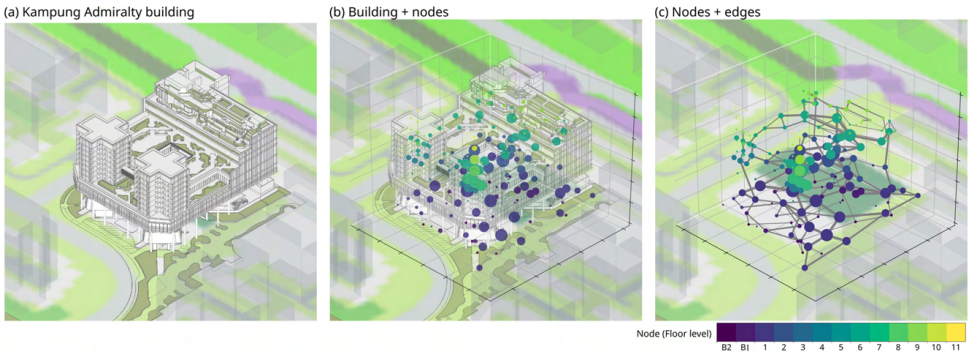


Figure 8. The study site. (**a**) The Kampung Admiralty building, (**b**) building and nodes, where size indicates incoming flows and color indicates floor level, (**c**) nodes and edges (weighted by the sum of flows in both directions).

Floor-group	*N*	*E*	Total flow	Node's flow range	Node's flow mean (std)	Edge's flow range	Edge's flow mean (std)
All nodes	165	476	4655	0–154	28.21 (± 37.03)	0–67	9.78 (± 13.47)
Car-parks	25	66	183	0–42	7.32 (± 9.75)	0–24	2.77 (± 5.12)
Level-1	33	102	1450	0–140	43.94 (± 37.27)	0–67	14.22 (± 14.35)
Level-2	19	54	741	0–144	39.00 (± 43.86)	0–64	13.72 (± 15.79)
Level-6*	25	72	878	4–131	35.12 (± 32.25)	0–45	12.19 (± 12.57)
Rooftop garden	19	46	77	0–11	4.05 (± 3.50)	0–6	1.67 (± 1.85)
Residential*	44	88	890	0–129	20.23 (± 32.20)	0–55	10.11 (± 15.62)
Other	9	16	61	0–40	6.78 (± 12.97)	0–40	3.81 (± 10.29)

Table 5. Basic information about KA's spatial adjacency network, including number of nodes (*N*), number of edges (*E*) within floor-group, and the mobility of nodes (incoming flow) and edge. *Level-6 and residential towers share several nodes on the sixth floor where they are interconnected, including the connection bridge and several public spaces. Category Other contains L3 (1 node), L4 (1 node), and L7 (7 nodes).

only of public, open, and freely accessible spaces as nodes. It is primarily built to analyze the relationship between public spaces and their usage based on the participant tracking data. Many closed or private spaces like shops, restaurants, medical facilities, etc., are not considered here as nodes. The exploded axonometric view of KA with functions in each floor and the abstract representation of nodes (and edges) in each floor are shown in Supplementary Figs. S11 and S12 respectively. The nodes are manually determined from the floor plans and the routing distance between them is measured from KA's floor plans. The network is constructed and analyzed through the NetworkX library in Python[69].

The aggregated movement flows are shown in Fig. 8b,c, in which the amount of traffic arriving at a node is shown by the node size, and the size of the edges measures the total edge weight irrespective of the flow direction.

Aggregated spatial networks

The movement data is also studied on an aggregated network. Two types of movement data are aggregated : (a) user movement from one location to its neighboring location in the network (4655 adjacent movements), (b) origin and destination of a trip taken by a user continuously moving with a pause time of less than 5 min in each location (361 trips). By aggregating the data by some combination of floor levels, we preserve the spatial adjacency and aim to study the movement flow across the vertical dimension. The floor-group components of the spatial network is also studied as sub-graphs without aggregation, thus preserving the individual location nodes and used to study within-group movement flows. The movement data is also aggregated by function that ignores the spatial adjacency and used to understand the movement flow between function categories.

Aggregation by floor

In this study, we adopted a specific framework to represent the multiple floors of this elevated building. Basements (B1 and B2) are grouped and defined as car-parks. Each of the levels L1, L2, and L6 became a group. Level-1 is the ground floor of KA, which is afforded a large open plaza space, several shops, and some entrances from different directions in the neighborhood. Level-2 is a food court, which can only be accessed by walking from L1 or the community lift connecting to the rest of KA's community side. Level-6 includes a childcare center, an active ageing hub, a playground, a fitness corner, a garden, and two connections to the social spaces in the residential tower side. Level-6 of the residential towers contains some public social spaces with two connections to the community side of KA that shares the space with the rest of L6. Therefore, the L6 floor and the residential towers share those public social spaces in the network analysis. Level-8 and L9 have a community farm, a community garden, and several sitting places; the two floors were grouped as a rooftop garden. Level-3 and L4 of KA is a medical center, which can be accessed by appointment only, hence containing only one node (lift lobby) for each of these two floors. In L7, the two community facilities (childcare center and active ageing hub) are connected to the lift lobbies, making seven nodes. These three levels (nine nodes) are grouped as 'other' and would not be discussed due to their simple structure. Some basic statistics of the network and the six floor groups (car-parks, L1, L2, L6, rooftop garden, and residential) are shown in Table 5.

Aggregation by function

The network is also aggregated by function (Table 6). A total of eight function categories are included: (a) entrance (12 nodes), (b) social (15 nodes), (c) commercial (24 nodes), (d) community (18 nodes), (e) garden (23 nodes), (f) residential (14 nodes), (g) vertical (48 nodes), and (h) corridor (11 nodes). The analyses and discussions in this study used both floor and function categories as the grouping methods.

Network measures

To understand the variation of outgoing and incoming flows, the outgoing/incoming entropy is commonly considered[70]:

Function	Locations
Entrance	Dropping points and walkways connecting to KA
Social	Seating area, community plaza, fitness corner
Commercial	Restaurant, bank, food court, supermarket, shops
Community	Childcare, elderly care, car-park
Garden	Garden around sitting area, walkways, and roof top gardens
Residential	Apartments
Vertical	Lift, escalator, and stairs lobby
Corridor	Corridor areas not associated with above

Table 6. Spatial functions.

$$\mathrm{H}^{\mathrm{out}}(c) = \frac{-\sum_{i=1}^{N} \mathrm{P}^{\mathrm{out}}(c,i) \times \log_2 \mathrm{P}^{\mathrm{out}}(c,i)}{\log_2 N}, \tag{1}$$

$$\mathrm{H}^{\mathrm{in}}(c) = \frac{-\sum_{i=1}^{N} \mathrm{P}^{\mathrm{in}}(i,c) \times \log_2 \mathrm{P}^{\mathrm{in}}(i,c)}{\log_2 N}, \tag{2}$$

where $\mathrm{P}^{\mathrm{out}}(c,i)$ is the proportion of flow leaving from group c to any group i divided by the total outflow from group c; $\mathrm{P}^{\mathrm{in}}(i,c)$ is the proportion of flow coming to group c from any group i divided by the total inflow to group c; N is the total number of groups. The term 'group' here refers to the floor groups or location categories.

Data Availability

The generated network and relevant materials are available at the following public GitHub repository (https://github.com/ajaymanivannan/Kampung_Admiralty_Dataset_v2), which includes the spatial network file with mobility data, anonymized participant's trip data, and beacon to zone transformation table. The research involving human subjects and the respective experiment protocol has been reviewed and approved by the Singapore University of Technology and Design (SUTD) Institutional Review Board (IRB) under reference code S-20-353. Informed consent has been obtained from all human participants and the experiment followed the necessary guidelines and regulations concerning research involving human subjects. The supplementary information file is available online along with this article.

Received: 29 July 2022; Accepted: 18 March 2024
Published online: 25 March 2024

References

1. United Nations. *The World's Cities in 2018-Data Booklet*. https://digitallibrary.un.org/record/3799524 (Accessed 22 February 2022) (2018).
2. Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913. https://doi.org/10.1038/467912a (2010).
3. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci.* **104**, 7301–7306. https://doi.org/10.1073/pnas.0610172104 (2007).
4. United Nations. *Resolution Adopted by the General Assembly on 6 July 2017, Work of the Statistical Commission Pertaining to the 2030 Agenda for Sustainable Development.*https://undocs.org/A/RES/71/313 *(Accessed 5 May 2022) (2017).*
5. House, E., O'Connor, C., Wolf, K., Israel, J. & Reynolds, T. *Outside Our Doors: The Benefits of Cities Where People and Nature Thrive*. https://www.nature.org/content/dam/tnc/nature/en/documents/Outside_Our_Doors_report.pdf (Accessed 6 June 2022) (2016).
6. Glaeser, E. L. & Kahn, M. E. The greenness of cities: Carbon dioxide emissions and urban development. *J. Urban Econ.* **67**, 404–418. https://doi.org/10.1016/j.jue.2009.11.006 (2010).
7. Meyer, W. B. *The Environmental Advantages of Cities: Countering Commonsense Antiurbanism* (The MIT Press, 2013).
8. Schröpfer, T. *Dense + Green Cities* (Birkhäuser, 2019).
9. Bibri, S. E., Krogstie, J. & Kärrholm, M. Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. *Dev. Built Environ.* **4**, 100021. https://doi.org/10.1016/j.dibe.2020.100021 (2020).
10. Mateo-Babiano, I. & Darchen, S. Vertical mixed use communities, a compact city model? In *6th State of Australian Cities Conference* (2013).
11. Bell, J. A mixed-use renaissance. *Mortgage Bank.* **64**, 66 (2004).
12. Rabianski, J., Gibler, K., Tidwell, O. A. & Clements, J. S. Mixed-use development: A call for research. *J. Real Estate Lit.* **17**, 205–230 (2009).
13. Freestone, R. Better planning and research for mixed-use developments. *Austral. Plann.* **45**, 14–15. https://doi.org/10.1080/07293682.2008.9982625 (2008).
14. Liao, Y., Gil, J., Pereira, R. H., Yeh, S. & Verendel, V. Disparities in travel times between car and transit: Spatiotemporal patterns in cities. *Sci. Rep.* **10**, 1–12 (2020).
15. Jemmali, M., Melhim, L. K. B., Alharbi, M. T., Bajahzar, A. & Omri, M. N. Smart-parking management algorithms in smart city. *Sci. Rep.* **12**, 1–15 (2022).
16. Batty, M. *et al.* Smart cities of the future. *Eur. Phys. J. Spl. Top.* **214**, 481–518. https://doi.org/10.1140/epjst/e2012-01703-3 (2012).
17. Al Dakheel, J., Del Pero, C., Aste, N. & Leonforte, F. Smart buildings features and key performance indicators: A review. *Sustain. Cities Soc.* **61**, 102328. https://doi.org/10.1016/j.scs.2020.102328 (2020).
18. Nembrini, J. & Lalanne, D. Human-building interaction: When the machine becomes a building. *In IFIP Conference on Human–Computer Interaction—INTERACT*, Vol. 348–369. https://doi.org/10.1007/978-3-319-67684-5_21 (Springer, 2017).

19. Simpson, M., Richter, K.-F., Wallgrün, J. O. & Klippel, A. Quantifying space, understanding minds: A visual summary approach. *J. Spatial Inf. Sci.* **1**, 95–136. https://doi.org/10.5311/JOSIS.2017.14.292 (2017).
20. Natapov, A., Kuliga, S., Dalton, R. C. & Hölscher, C. Linking building-circulation typology and wayfinding: Design, spatial analysis, and anticipated wayfinding difficulty of circulation types. *Archit. Sci. Rev.* **63**, 34–46. https://doi.org/10.1080/00038628.2019.1675041 (2020).
21. Rahimi, M., Malek, M. R., Claramunt, C. & Le Pors, T. A topology-based graph data model for indoor spatial-social networking. *Int. J. Geogr. Inf. Sci.* **35**, 2517–2539. https://doi.org/10.1080/13658816.2021.1912349 (2021).
22. Ericson, J. D., Chrastil, E. R. & Warren, W. H. Space syntax visibility graph analysis is not robust to changes in spatial and temporal resolution. *Environ. Plann. Urban B Anal. City Sci.* **48**, 1478–1494. https://doi.org/10.1177/2399808319897624 (2021).
23. Mashhadi, A. et al. Exploring space syntax on entrepreneurial opportunities with Wi-Fi analytics. In *Proc. 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* 658–669. https://doi.org/10.1145/2971648.2971745 (2016).
24. Netto, V. M. 'What is space syntax not?' Reflections on space syntax as sociospatial theory. *Urban Des. Int.* **21**, 25–40. https://doi.org/10.1057/udi.2015.21 (2016).
25. Karimi, K. A configurational approach to analytical urban design: 'Space syntax' methodology. *Urban Des. Int.* **17**, 297–318. https://doi.org/10.1057/udi.2012.19 (2012).
26. Ravulaparthy, S. K. & Goulias, K. G. Characterizing the composition of economic activities in central locations: Graph-theoretic approach to urban network analysis. *Transp. Res. Rec.* **2430**, 95–104. https://doi.org/10.3141/2430-10 (2014).
27. Porta, S., Crucitti, P. & Latora, V. The network analysis of urban streets: A primal approach. *Environ. Plann. B Plann. Des.* **33**, 705–725. https://doi.org/10.1068/b32045 (2006).
28. Batty, M. *The New Science of Cities* (The MIT Press, 2013).
29. Barabási, A.-L. & Pósfai, M. *Network Science* (Cambridge University Press, 2016).
30. Chang, D. Spatial choice and preference in multilevel movement networks. *Environ. Behav.* **34**, 582–615. https://doi.org/10.1177/0013916502034005002 (2002).
31. Tunçer, B. & Benita, F. Data-driven thinking for measuring the human experience in the built environment. *Int. J. Archit. Comput.* **1**, 14780771211025142. https://doi.org/10.1177/14780771211025142 (2021).
32. Pont, M. B. & Marcus, L. What can typology explain that configuration cannot. In *Proc. 10th International Space Syntax Symposium*, Vol. 43, 1–16 (2015).
33. Willis, A., Gjersoe, N., Havard, C., Kerridge, J. & Kukla, R. Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments. *Environ. Plann. B Plann. Des.* **31**, 805–828. https://doi.org/10.1068/b3060 (2004).
34. Omer, I. & Kaplan, N. Using space syntax and agent-based approaches for modeling pedestrian volume at the urban scale. *Comput. Environ. Urban Syst.* **64**, 57–67. https://doi.org/10.1016/j.compenvurbsys.2017.01.007 (2017).
35. Rout, A. & Willett, W. (Big) data in urban design practice: Supporting high-level design tasks using a visualization of human movement data from smartphones. In *Urban Informatics and Future Cities* (eds Geertman, S. C. M. *et al.*) 301–318 (Springer, 2021).
36. Boeing, G. D. *Methods and Measures for Analyzing Complex Street Networks and Urban Form*. Ph.D. thesis, University of California (2017).
37. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101. https://doi.org/10.1016/j.physrep.2010.11.002 (2011).
38. Tabak, V., de Vries, B. & Dijkstra, J. Simulation and validation of human movement in building spaces. *Environ. Plann. B Plann. Des.* **37**, 592–609. https://doi.org/10.1068/b35127 (2010).
39. Manivannan, A., Chin, W. C. B., Barrat, A. & Bouffanais, R. On the challenges and potential of using barometric sensors to track human activity. *Sensors* **20**, 6786. https://doi.org/10.3390/s20236786 (2020).
40. Manivannan, A. *et al.* A framework for the identification of human vertical displacement activity based on multi-sensor data. *IEEE Sens. J.* **22**, 8011–8029. https://doi.org/10.1109/JSEN.2022.3157806 (2022).
41. Newman, M. *Networks* (Oxford University Press, 2018).
42. Srikanth, A. D. S., Chin, W. C. B., Bouffanais, R. & Schroepfer, T. Complexity science for urban solutions. In *Artificial Intelligence in Urban Planning and Design: Technologies, Implementation, and Impacts* (eds As, I. *et al.*) 39–58 (Elsevier, 2022).
43. Zhang, L. & Chiaradia, A. J. Urban volumetrics: Spatial complexity and wayfinding, extending space syntax to three dimensional space. Preprint at http://arxiv.org/abs/2012.14419 (2020).
44. Lai, Y. & Kontokosta, C. E. Quantifying place: Analyzing the drivers of pedestrian activity in dense urban environments. *Landsc. Urban Plann.* **180**, 166–178 (2018).
45. van Meeteren, M. About being in the middle: Conceptions, models and theories of centrality in urban studies. In *Handbook of Cities and Networks* (eds Neal, Z. P. & Rozenblat, C.) 252–271 (Edward Elgar Publishing, 2021).
46. Zuniga-Teran, A. A. *et al.* Designing healthy communities: Testing the walkability model. *Front. Archit. Res.* **6**, 63–73. https://doi.org/10.1016/j.foar.2016.11.005 (2017).
47. Wang, W., Osaragi, T. & Tagashira, M. Sequential patterns of daily human activity extracted from person trip survey data. In *Urban Informatics and Future Cities* (eds Geertman, S. C. M. *et al.*) 257–275 (Springer, 2021).
48. Serra, M. & Hillier, B. Angular and metric distance in road network analysis: A nationwide correlation study. *Comput. Environ. Urban Syst.* **74**, 194–207 (2019).
49. Zhang, L., Chiaradia, A. & Zhuang, Y. A configurational accessibility study of road and metro network in Shanghai, China. In *Recent Developments in Chinese Urban Planning: Selected Papers from the 8th International Association for China Planning Conference, Guangzhou, China, June 21–22, 2014* 219–245 (Springer, 2015).
50. Zhang, L. & Chiaradia, A. From axial to pedestrian path-centre line: The case of 3d pedestrian network in Hong Kong, central. In *The 12th International Space Syntax Symposium (12SSS)* (Beijing JiaoTong University, 2019).
51. Cooper, C. H. Spatial localization of closeness and betweenness measures: A self-contradictory but useful form of network analysis. *Int. J. Geogr. Inf. Sci.* **29**, 1293–1309 (2015).
52. WOHA. *Kampung Admiralty/WOHA*. https://www.archdaily.com/904646/kampung-admiralty-woha (Accessed 14 June 2021) (2018).
53. Katherine, A. *WOHA's Kampung Admiralty Singapore Named 2018 Building of the Year at World Architecture Festival*. https://www.archdaily.com/906940/wohas-kampung-admiralty-singapore-named-2018-building-of-the-year-at-world-architecture-festival (Accessed 14 June 2021) (2018).
54. Block, I. *WOHA Creates Green Community with Kampung Admiralty in Singapore*. https://www.dezeen.com/2018/12/07/kampung-admiralty-woha-singapore-world-building-year/ (Accessed 14 June 2021) (2018).
55. Jiang, B. Ranking spaces for predicting human movement in an urban environment. *Int. J. Geogr. Inf. Sci.* **23**, 823–837. https://doi.org/10.1080/13658810802022822 (2009).
56. Chin, W. C. B. & Wen, T.-H. Geographically modified PageRank algorithms: Identifying the spatial concentration of human movement in a geospatial network. *PLoS ONE* **10**, 1–23. https://doi.org/10.1371/journal.pone.0139509 (2015).
57. Ducruet, C., Lee, S.-W. & Ng, A. K. Centrality and vulnerability in liner shipping networks: Revisiting the northeast Asian port hierarchy. *Maritime Policy Manag.* **37**, 17–36. https://doi.org/10.1080/03088830903461175 (2010).
58. Guimera, R., Mossa, S., Turtschi, A. & Amaral, L. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci.* **102**, 7794–7799. https://doi.org/10.1073/pnas.0407994102 (2005).

59. Fu, M., Liu, R. & Hon, C. K. Walkability evaluation of building circulation based on user preference. *Eng. Constr. Archit. Manag.* **28**, 2904–2924. https://doi.org/10.1108/ECAM-06-2020-0398 (2020).
60. Pappalardo, L., Manley, E., Sekara, V. & Alessandretti, L. Future directions in human mobility science. *Nat. Comput. Sci.* **1**, 1–13 (2023).
61. Barbosa, H. *et al.* Human mobility: Models and applications. *Phys. Rep.* **734**, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001 (2018).
62. Srikanth, A. D. S., Chin, W. C. B., Bouffanais, R. & Schroepfer, T. Complexity science-based spatial performance analyses of UNStudio/DP Architects' SUTD campus and WOHA's Kampung Admiralty. In *Artificial Intelligence in Urban Planning and Design: Technologies, Implementation, and Impacts* (eds As, I. *et al.*) 217–244 (Elsevier, 2022).
63. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10**, e0136497 (2015).
64. Manivannan, A., Yow, W. Q., Bouffanais, R. & Barrat, A. Are the different layers of a social network conveying the same information? *EPJ Data Sci.* **7**, 1–26 (2018).
65. Alessandretti, L., Lehmann, S. & Baronchelli, A. Understanding the interplay between social and spatial behaviour. *EPJ Data Sci.* **7**, 36 (2018).
66. Liang, X. & Kang, Y. A review of spatial network insights and methods in the context of planning: Applications, challenges, and opportunities. In *Urban Informatics and Future Cities* (eds Geertman, S. C. M. *et al.*) 71–91 (Springer, 2021).
67. Alessandretti, L., Sapiezynski, P., Lehmann, S. & Baronchelli, A. Multi-scale spatio-temporal analysis of human mobility. *PLoS ONE* **12**, e0171686 (2017).
68. Housing & Development Board, Singapore. *Kampung Admiralty*. https://www.hdb.gov.sg/residential/where2shop/explore/woodlands/kampung-admiralty (Accessed 28 April 2022) (2021).
69. Hagberg, A., Swart, P. & Chult, S. D. *Exploring Network Structure, Dynamics, and Function Using Networkx. Tech Rep.* (Los Alamos National Lab, 2008).
70. Chin, W. C. B. & Bouffanais, R. Spatial super-spreaders and super-susceptibles in human movement networks. *Sci. Rep.* **10**, 1–19. https://doi.org/10.1038/s41598-020-75697-z (2020).

Acknowledgements
The project is funded by the Ministry of National Development (MND) Singapore affiliated with the project—Vertical Cities: Complex Emergent Patterns of Space Use and in High-Density Urban Contexts (Kampung Admiralty pilot project). The authors thank Dr. BT Balamurali for the conversations on data collection and machine learning, and Anjanaa Devi Srikanth and Chirag Hablani for their assistance with the data collection.

Author contributions
T.S. and R.B. conceived the experiments. S.G., A.M., W.B., and D.W. conducted the experiments. A.M. and W.B analyzed the results. W.B, A.M. and R.B. wrote the manuscript. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-57475-3.

Correspondence and requests for materials should be addressed to R.B.

Supplementary Information: Effects of the interplay between topology and function of an integrated urban development on patterns of user movement

Ajaykumar Manivannan[1,+], Wei Chien Benny Chin[2,+], Srilalitha Gopalakrishnan[3], Daniel KH Wong[4], Thomas Schroepfer[3,4], and Roland Bouffanais[5,*]

[1]Department of Mechanical Engineering, University of Ottawa, Ottawa, K1N 6N5, Canada
[2]Department of Geography, National University of Singapore, 117568, Singapore
[3]Future Cities Laboratory Global, Singapore-ETH Centre, 138602, Singapore
[4]Architecture and Sustainable Design, Singapore University of Technology and Design, 487372, Singapore
[5]Department of Computer Science & Global Studies Institute, University of Geneva, CH-1211 Genève, Switzerland
[*]roland.bouffanais@unige.ch
[+]these authors contributed equally to this work

1 Removing Vertical node in OD matrix

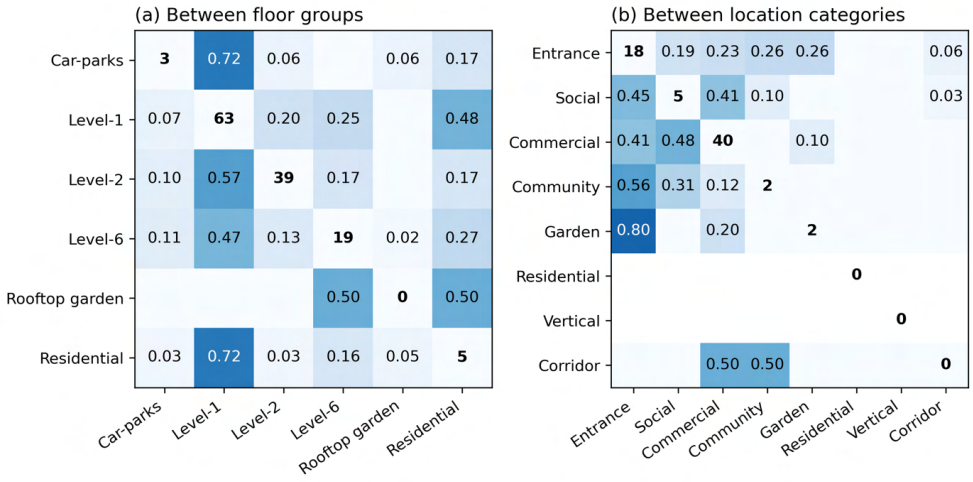


Figure S1. Row normalized origin-destination matrix — after removing vertical node as origin or destination. (a) Aggregated floor-groups network, and (b) aggregated function categories network. The row normalized values indicated the proportion of flows from one origin to every destination.

2 Effect of Pause time of 10 minutes

Table S1. Pearson's correlation results between topological network centralities and mobility (incoming flow) for all and each floor-group. The closeness and betweenness centralities are weighted by routing distance. Pause time = 10 minutes.

Floor-groups	In-degree	Closeness (weighted)	Betweenness (weighted)
All nodes	0.466***	0.420***	0.390***
Car-parks	-0.202	0.105	-0.005
Level-1	0.551**	0.420*	0.644***
Level-2	0.632**	0.812***	0.765***
Level-6	0.431*	0.22	0.24
Rooftop garden	0.022	0.504	0.637*
Residential	0.380*	0.364	0.406*

*** : p-value ≤ 0.001; ** : p-value ≤ 0.01; * : p-value ≤ 0.05.

Table S2. The proportion of internal flows and external flows for floor-groups and function categories. 'Residential' in the Floor-group refers to both residential towers, while 'Residential' in the function categories denotes the individual apartments of the two residential towers that exclude vertical transportation and corridors. Pause time = 10 minutes.

Grouping	Category	Total	Internal	External
Floor-group	Car-parks	190	0.75	0.25
Floor-group	Level-1	1479	0.89	0.11
Floor-group	Level-2	679	0.82	0.18
Floor-group	Level-6	571	0.86	0.14
Floor-group	Rooftop garden	44	0.89	0.11
Floor-group	Residential	769	0.89	0.11
Function category	Entrance	372	0.49	0.51
Function category	Social	503	0.34	0.66
Function category	Commercial	1028	0.78	0.22
Function category	Community	284	0.17	0.83
Function category	Garden	233	0.32	0.68
Function category	Residential	36	0	1
Function category	Vertical	1225	0.71	0.29
Function category	Corridor	171	0.01	0.99

Table S3. Normalized outgoing entropy (H^{out}) and incoming entropy (H^{in}) of the floor-group and function categories based on origin and destination of a path. Pause time = 10 minutes.

Floor-group	H^{out}	H^{in}	Function category	H^{out}	H^{in}
Car-parks	0.494	0.526	Entrance	0.722	0.701
Level-1	0.692	0.704	Social	0.582	0.677
Level-2	0.465	0.535	Commercial	0.644	0.612
Level-6	0.688	0.618	Community	0.617	0.569
Rooftop garden	0	0.613	Garden	0.542	0.445
Residential	0.412	0.484	Residential	0	0
			Vertical	0.671	0.609
			Corridor	0	0

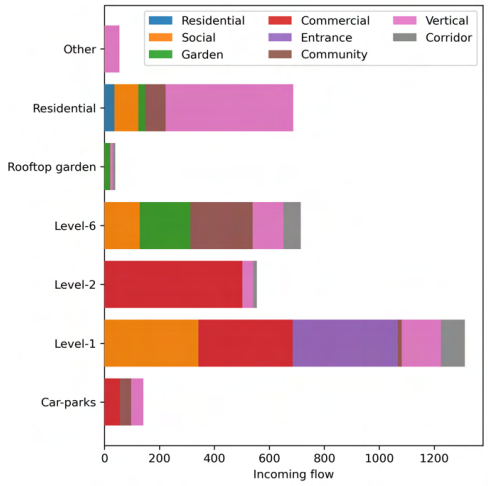


Figure S2. Distribution of inflow by floor-group and program categories. Pause time = 10 minutes.

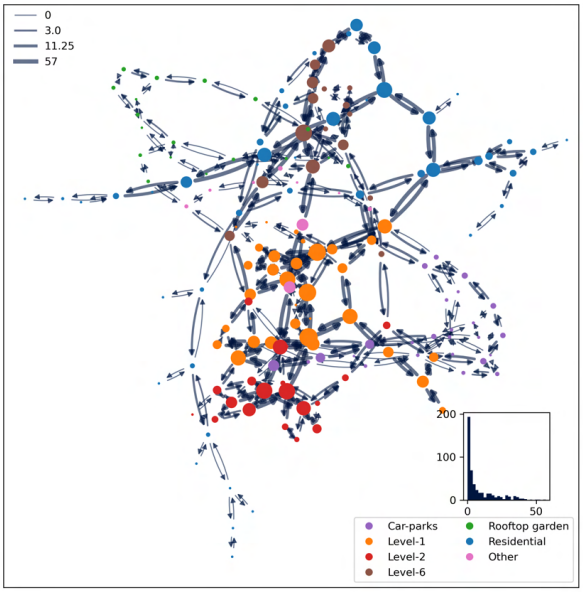


Figure S3. KA spatial network. Nodes are colored by floor group and sized by movement inflow. Directed edges are sized by the number of movement flows in respective directions. Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, 75^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization). Pause time = 10 minutes.

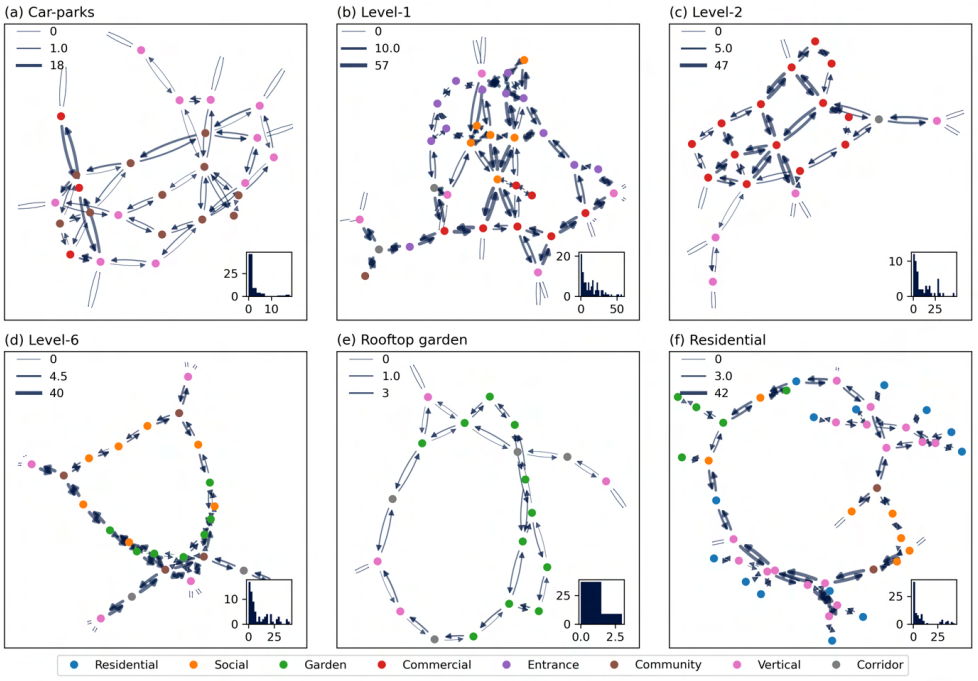


Figure S4. Subgraphs of the spatial network. (a) Car-parks (B1 and B2), (b) L1, (c) L2 (food court), (d) L6, (e) rooftop garden (L8 and L9), and (f) the two residential towers. The colors indicate program categories. Bottom right plot: Histogram of flow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization). Pause time = 10 minutes.

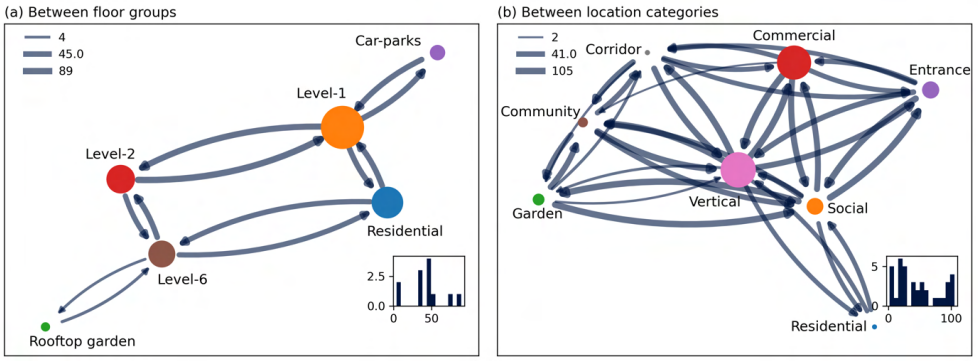


Figure S5. Networks of aggregated flow. (a) By floor-groups and (b) by function categories. The node sizes indicate internal flows (self-loop). Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization). Pause time = 10 minutes.

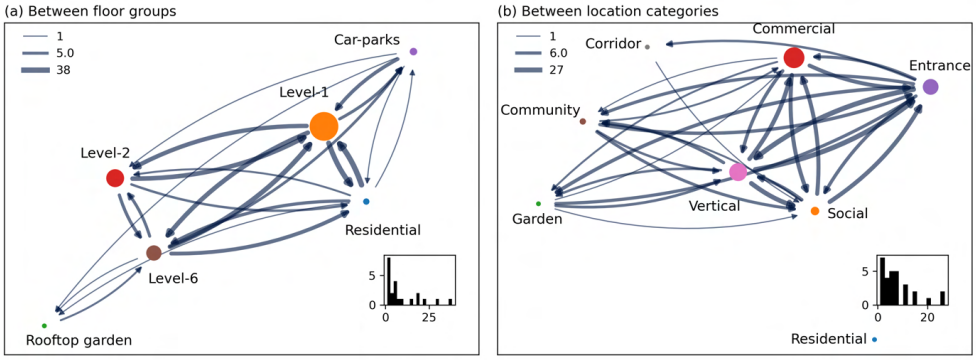


Figure S6. Networks of aggregated flow based on origin and destination of a path (a) By floor-groups and (b) by function categories. The node sizes indicate internal flows (self-loop). Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization). Pause time = 10 minutes.

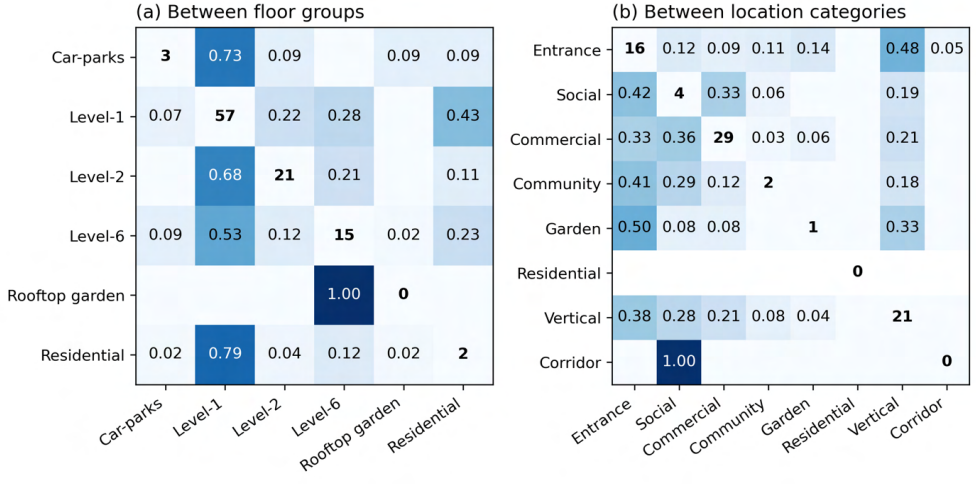


Figure S7. Row normalized origin-destination matrix. (a) Aggregated floor-groups network, and (b) aggregated function categories network. The row normalized values indicated the proportion of flows from one origin to every destination. Pause time = 10 minutes.

Table S4. Basic information about KA's spatial adjacency network, including number of nodes (N), number of edges (E) within floor-group, and the mobility of nodes (incoming flow) and edge. Pause time = 10 minutes.

Floor-group	N	E	Total flow	Node's flow range	Node's flow mean (std)	Edge's flow range	Edge's flow mean (std)
All nodes	165	476	3852	0–133	23.35 (±30.67)	0–57	8.09 (±11.19)
Car-parks	25	66	142	0–32	5.68 (±7.35)	0–18	2.15 (±3.89)
Level-1	33	102	1312	0–120	39.76 (±33.39)	0–57	12.86 (±12.87)
Level-2	19	54	555	0–103	29.21 (±30.65)	0–47	10.28 (±11.07)
Level-6*	25	72	715	4–110	28.60 (±26.89)	0–40	9.93 (±10.75)
Rooftop garden	19	46	39	0–4	2.05 (±1.57)	0–3	0.85 (±0.98)
Residential*	44	88	687	0–98	15.61 (±24.13)	0–42	7.81 (±11.76)
Other	9	16	55	0 – 37	6.11 (±11.84)	0 – 37	3.44 (±9.38)

*Level-6 and residential towers share several nodes at the sixth floor where they are interconnected, including the connection bridge and several public spaces. Category Other contains L3 (1 node), L4 (1 node), and L7 (7 nodes).

3 Weekday Vs Weekend

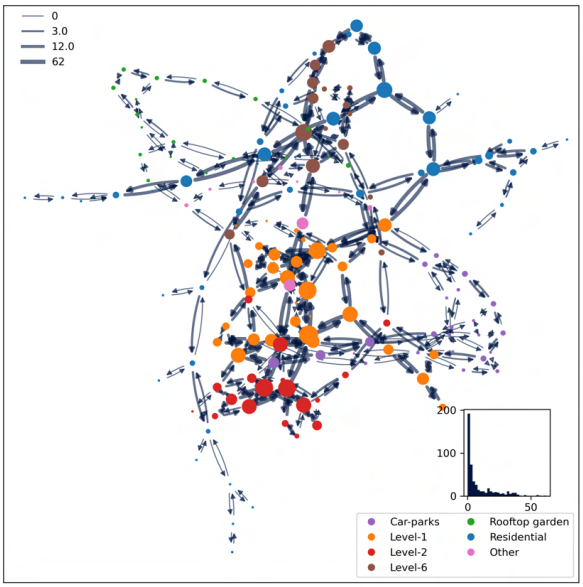


Figure S8. KA spatial network—Weekday. Nodes are colored by floor group and sized by movement inflow. Directed edges are sized by the number of movement flows in respective directions. Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, 75^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

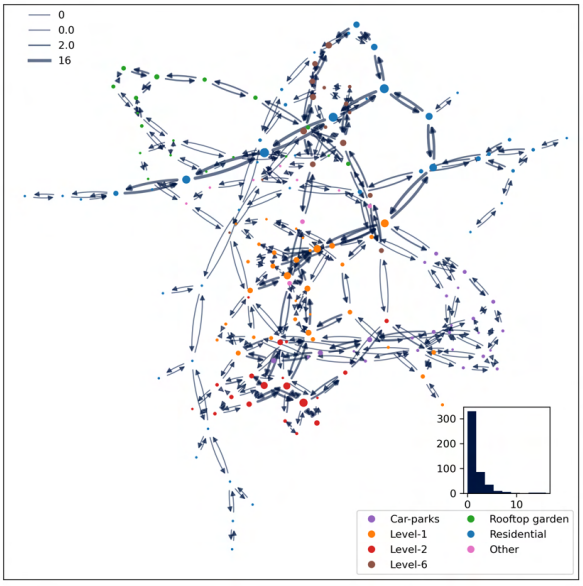


Figure S9. KA spatial network—Weekend. Nodes are colored by floor group and sized by movement inflow. Directed edges are sized by the number of movement flows in respective directions. Bottom right plot: Histogram of inflow count. Top left: Legend indicating flow size for minimum, 50^{th} percentile, 75^{th} percentile, and maximum values (edges sized by taking the logarithm of flow count for better contrast in visualization).

Table S5. Weekday: Basic information about KA's spatial adjacency network, including the number of nodes (N), number of edges (E) within floor-group, and the mobility of nodes (incoming flow) and edge.

Floor-group	N	E	Total flow	Node's flow range	Node's flow mean (std)	Edge's flow range	Edge's flow mean (std)
All nodes	165	476	3987	0–144	24.16 (±32.35)	0–62	8.38 (±11.70)
Car-parks	25	66	144	0–35	5.76 (±8.15)	0–20	2.18 (±4.27)
Level-1	33	102	1318	0–130	39.94 (±33.75)	0–62	12.92 (±13.17)
Level-2	19	54	645	0–128	33.95 (±38.38)	0–57	11.94 (±13.84)
Level-6*	25	72	747	4–117	29.88 (±28.34)	0–41	10.38 (±11.18)
Rooftop garden	19	46	42	0–5	2.21 (±1.73)	0–4	0.91 (±1.06)
Residential*	44	88	704	0–101	16.00 (±24.51)	0–39	8.00 (±11.84)
Other	9	16	56	0 – 37	6.22 (±11.92)	0 – 37	3.50 (±9.46)

*Level-6 and residential towers share several nodes on the sixth floor where they are interconnected, including the connection bridge and several public spaces. Category Other contains L3 (1 node), L4 (1 node), and L7 (7 nodes).

Table S6. Weekend: Basic information about KA's spatial adjacency network, including the number of nodes (N), number of edges (E) within floor-group, and the mobility of nodes (incoming flow) and edge.

Floor-group	N	E	Total flow	Node's flow range	Node's flow mean (std)	Edge's flow range	Edge's flow mean (std)
All nodes	165	476	668	0–29	4.05 (±5.89)	0–16	1.40 (±2.35)
Car-parks	25	66	39	0–7	1.56 (±1.70)	0–4	0.59 (±0.95)
Level-1	33	102	132	0–18	4.00 (±4.56)	0–7	1.29 (±1.70)
Level-2	19	54	96	0–23	5.05 (±6.09)	0–8	1.78 (±2.30)
Level-6*	25	72	131	0–14	5.24 (±4.71)	0–7	1.82 (±1.97)
Rooftop garden	19	46	35	0–7	1.84 (±2.18)	0–3	0.76 (±1.03)
Residential*	44	88	186	0–29	4.23 (±8.00)	0–16	2.11 (±3.99)
Other	9	16	5	0 – 3	0.56 (±1.07)	0 – 3	0.31 (±0.85)

*Level-6 and residential towers share several nodes at the sixth floor where they are interconnected, including the connection bridge and several public spaces. Category Other contains L3 (1 node), L4 (1 node), and L7 (7 nodes).

4 Similarity in mobility flow between different days of a user

The *cosine similarity* for each node i, which measures the similarity between this node and its neighborhoods in the two networks. If $w_{ij,1}$ and $w_{ij,2}$ denote the weights on the links from i to j respectively in networks 1 and 2, the cosine similarity of i is defined as

$$\mathrm{sim}_{1,2}(i) = \frac{\sum_j w_{ij,1} w_{ij,2}}{\sqrt{\sum_j w_{ij,1}^2}\sqrt{\sum_j w_{ij,2}^2}}. \tag{1}$$

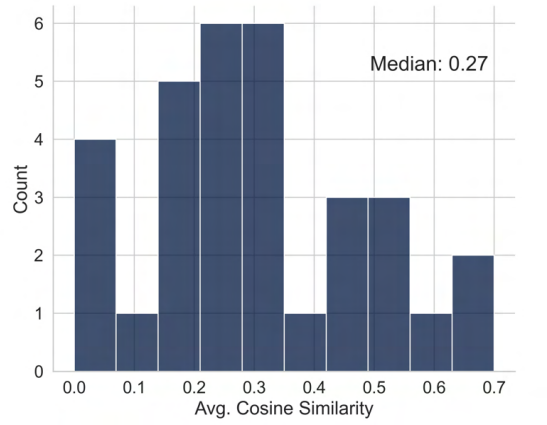


Figure S10. Histogram of Average Cosine similarity between 1-day aggregated networks. It is calculated for different days of a user for all users. Average Cosine Similarity (CS) for user i = (Sum of CS of all pairs of 1-day network for user i)/number of pairs of 1-day network for user i

5 Floor plan to KA spatial network

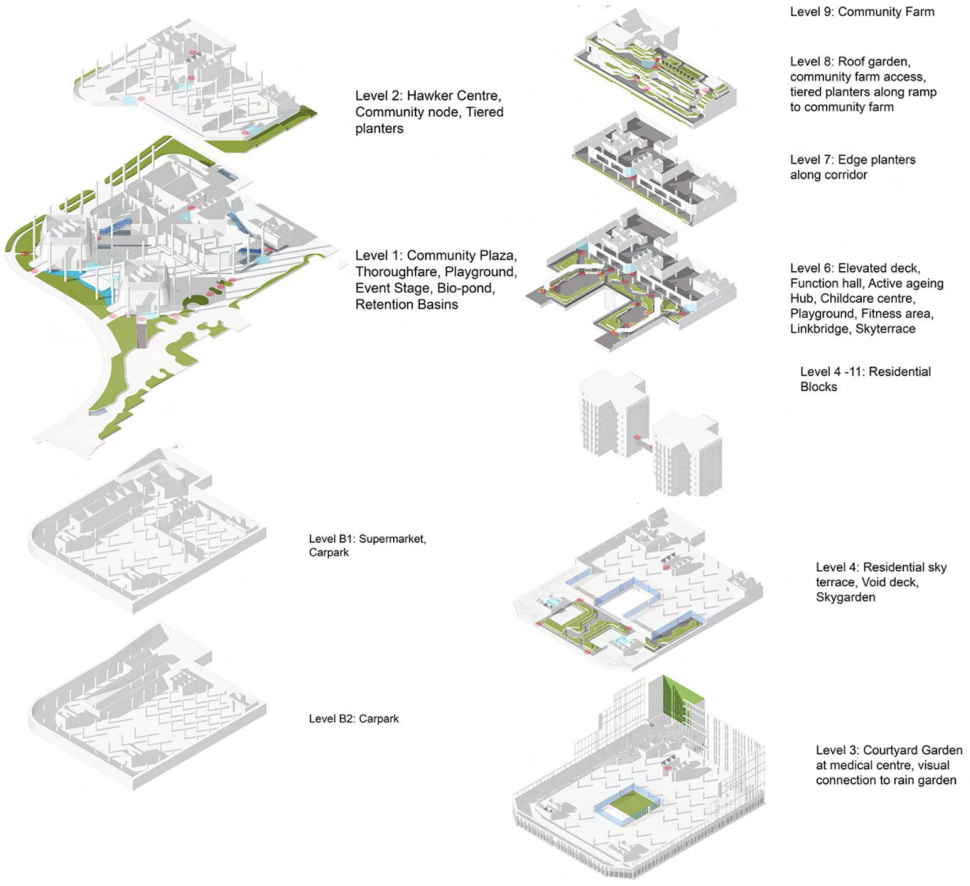
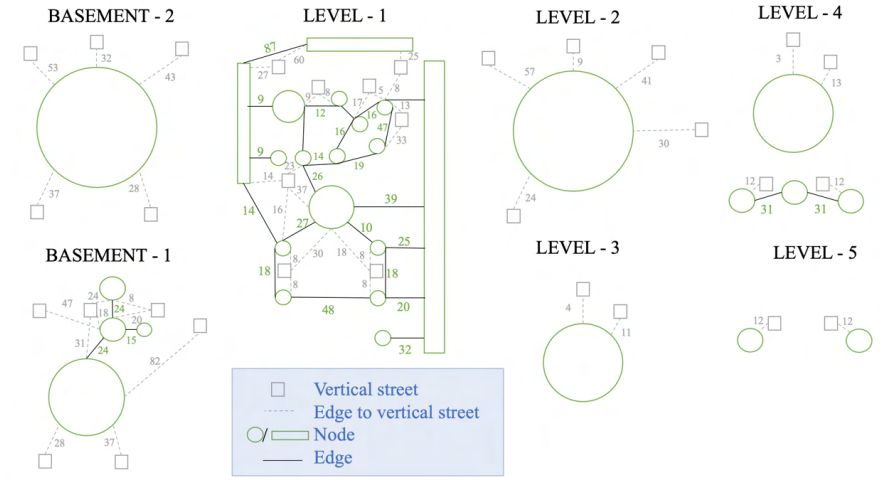
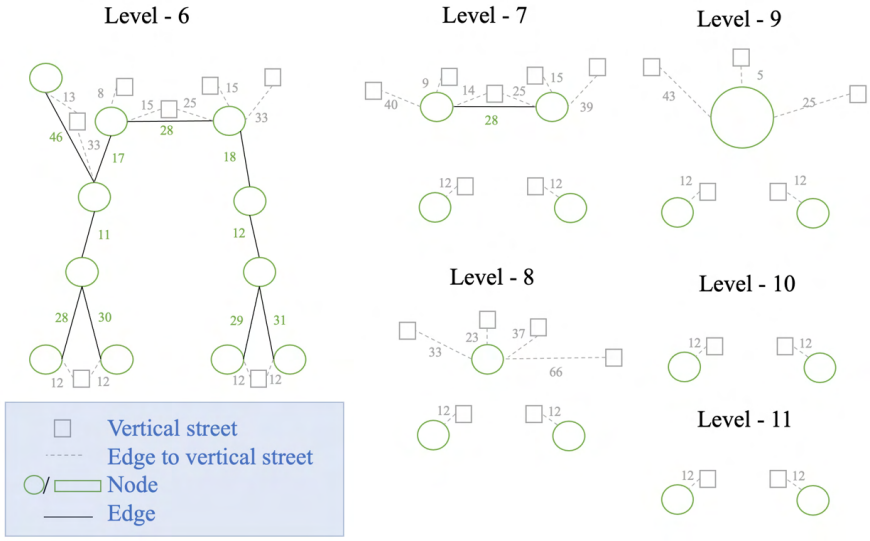


Figure S11. Exploded Axonometric of Kampung Admiralty.

Figure S12. Abstract representation of floor plan to network transformation by each floor. The number near the edge represents the routing distance in meters between nodes. The distance is calculated from the floor plans provided by the Architects of KA. Nodes from each floor are stacked on top of each other and connected through vertical facilities like lifts, stairs, and escalators.



(a)



(b)

6 Nodes in KA spatial network

A node in the network is an approximate and variable area of space in the KA building that serves a particular function. For example, a walkway followed by an entrance, and a lift area will be divided into three nodes Walkway, Entrance, and Lift (Figure S13). An exception to this case is when the space is large enough that it requires sub-divisions for more accurate tracking of a user along the floor. For example, in Level 1, the Community Plaza (Figure S13) is divided into four nodes (e.g., CP_Zone_N where $N = 1, 2, 3,$ and 4). This allows us to calculate routing distance more accurately along the spatial network. A routing distance is the distance between the centroids of the two connected nodes.

The beacon locations are not always related to a particular node. While we make all steps to confine a beacon to capture one particular node (by adjusting the transmission range and orientation of the beacon), this is not always possible or accurate. For these reasons, we used Machine Learning models to relate beacons to nodes based on the trained data that contains records (Beacon signature) of beacons when passing through each node.

There are 165 nodes and 124 beacons. The disparity between the number of nodes and beacons is because many nodes (such as corridors or walkways) do not require a dedicated beacon, as they may serve as an interconnecting node between two other nodes that are tracked by a beacon. Hence not all nodes require a beacon and its location in a user's path is implicitly assumed.

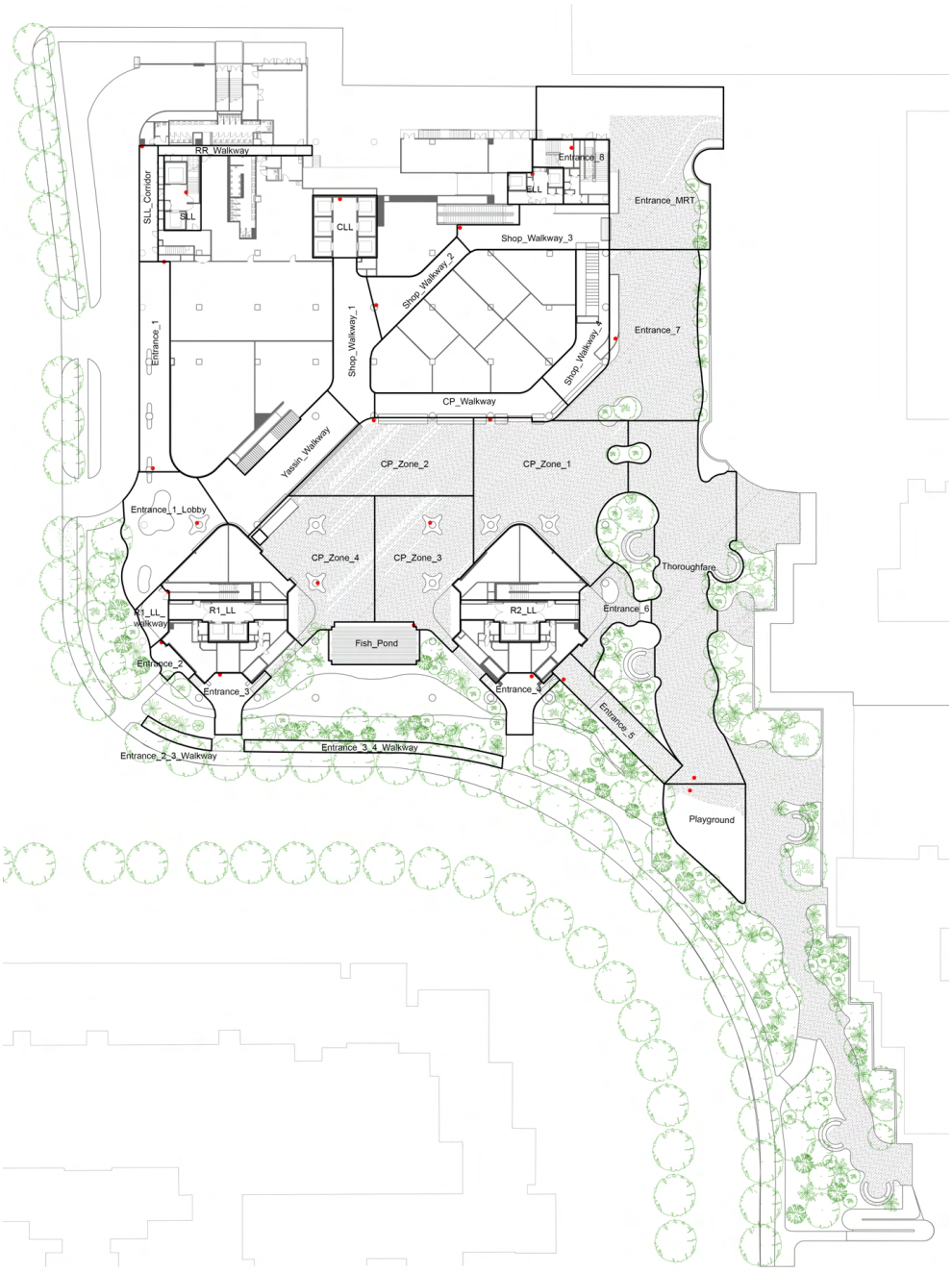


Figure S13. KA Level 1 node divisions

7 Demographic details based on mobility data

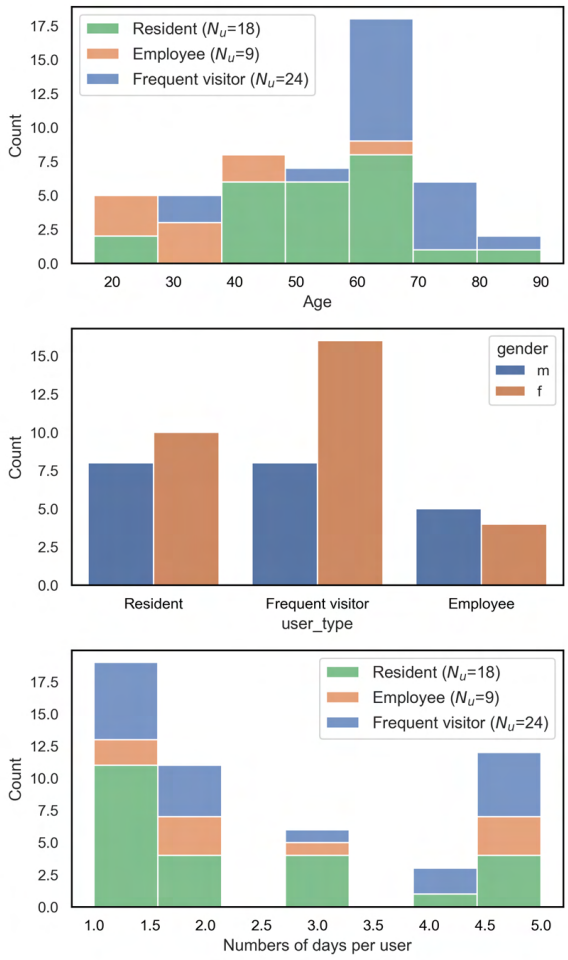


Figure S14. Demographic details by category of participants. Number of users per (a) age, (b) gender, and (c) number of days

Part V

Discussion and Conclusions

8

Discussion

This thesis set out to advance our understanding of human mobility by incorporating a critical yet previously underexplored dimension: vertical displacement in urban environments. The overarching goal was to bridge the knowledge gaps related to vertical mobility, develop methodologies to track and characterize it at scale, and examine how it interacts with horizontal movements and built environment features. Through four interrelated objectives, we addressed (1) the behavior and interpretation of barometric pressure as a primary sensor for the study and tracking of vertical movement, (2) the development of a machine learning framework to accurately infer vertical displacement from wearable sensor data, (3) the large-scale characterization of vertical human movement and its interplay with horizontal mobility, and (4) the quantitative analysis of how topology and spatial function jointly influence movement patterns in integrated vertical developments. In essence, the key findings indicate that vertical mobility is both prevalent and complex, and it must be considered to achieve a complete understanding of urban movement dynamics.

8.1 Key Findings and Their Interpretation

Finding 1 (RO-1: Understanding Barometric Pressure Behavior): A crucial initial step was to characterize the factors influencing barometric pressure measurements and their suitability for tracking human vertical activity. This research found that barometric pressure is sensitive not only to altitude changes but also to atmospheric conditions, building pressurization, sensor drift, and transient conditions created by human movement and the built infrastructure. Compared to previous studies that had not fully accounted for these numerous confounding factors, our experiments demonstrated the magnitude and temporal scales at which these influences operate. Interpreting these findings in light of the literature on atmospheric and indoor environmental physics suggests that vertical displacement can indeed be measured reliably with barometric sensors only if one accounts for factors beyond simple elevation changes. Thus, barometer-based vertical sensing can be integrated into broader sensing platforms, improving indoor positioning, health monitoring, and navigation technologies. To further bolster these insights, future work could delve deeper into long-term sensor stability analyses and explore more robust calibration strategies under changing climatic conditions.

Finding 2 (RO-2: Developing a Framework for Extracting Vertical Displacement): Building on the understanding established from RO-1, we developed a machine learning framework capable of accurately detecting vertical displacement (e.g., elevator rides, stair climbing) from wearable multi-sensor data. Unlike previous studies limited to controlled environments or small samples, our approach was tested at a city scale with thousands of participants, achieving high accuracy and F_1-scores. Interpreted alongside earlier human activity recognition research, the high classification performance not only validates the framework's robustness but also expands its applicability to diverse urban environments. Importantly, the success of this method highlights that vertical movement is both measurable and generalizable, even under real-world complexity. The implications for practice are substantial: improved indoor navigation, enhanced energy expenditure estimation, and more accurate modeling of human mobility dynamics. There is room to expand on this framework by incorporating emerging sensor modalities (e.g., low-power Bluetooth, lidar), refining context-aware algorithms, or investigating transferability across cities and cultures.

Finding 3 (RO-3: Uncovering Vertical Mobility Patterns and

Their Interplay with Horizontal Movements): Having established reliable vertical displacement detection methods, we applied them to large-scale human data and uncovered that vertical mobility is pervasive, non-trivial, and tightly intertwined with horizontal movement. We observed heavy-tailed distributions of vertical displacement distances and waiting times, much as horizontal mobility studies have documented for planar travel. Notably, while vertical travel distances are comparatively shorter than horizontal ones, the complexity and timing patterns closely mirror established daily travel rhythms. This suggests that vertical movements, though often hidden, follow similar underlying mechanisms and constraints as horizontal travel.

Drawing parallels with existing human mobility literature, which often focuses on ground-level or planar analyses, our findings reveal that ignoring the vertical dimension can omit an essential layer of urban complexity. The interplay between vertical and horizontal mobility is increasingly relevant as cities grow upward. Planners, engineers, and policymakers should recognize these three-dimensional patterns to design efficient, livable, and energy-saving vertical infrastructure. To deepen these insights, additional data sources—such as building occupancy records, elevator usage logs, or inter-building walkway data—could refine our understanding of how vertical connectivity reshapes urban flows.

Finding 4 (RO-4: Quantifying the Influence of Topology and Function in Integrated Developments): In the final phase, the thesis examined how topology (network structure) and spatial function (the purpose and use of spaces) affect user movement within a three-dimensional integrated development. Our network-theoretic analysis of Kampung Admiralty, a state-of-the-art vertical mixed-use development, showed a moderate yet consistent correlation (40%) between node centrality metrics and movement flows. While previous city-scale analyses also found correlations between connectivity and movement, the integration of vertical corridors, multi-level plazas, and "vertical streets" introduces a new layer of complexity. Functionally, the placement of amenities and services within a vertically connected environment strongly influences route choices, patterns of use, and circulation intensity.

These findings advance the field of data-driven architectural and urban-planning design and space syntax studies, confirming that vertical mobility patterns are shaped by both the underlying spatial graph and the purposeful allocation of functionalities within that space. Expanding on these insights, future work could incorporate user preferences, demographic profiles, or so-

cial interaction networks to better understand how different groups navigate complex, multi-level built environments.

8.2 Broader Implications

Theoretical contribution: The integration of vertical mobility into human movement studies extends the theoretical foundations of urban complexity, human mobility modeling, and transportation science. By empirically demonstrating that vertical and horizontal mobility share fundamental scaling properties, complexity, and temporal rhythms, this thesis helps refine existing models of human movement. Instead of treating cities as flat planes, we introduce a volumetric perspective, pushing future theories toward three-dimensional frameworks.

Practical Applications: Practically, these findings inform several domains:

- Urban Planning and Design: Architects and planners can leverage vertical mobility insights to improve building layouts, place amenities at accessible vertical nodes, and design user-friendly, resource-efficient high-rise infrastructures.

- Transportation and Navigation Systems: Understanding when, where, and how vertical mobility occurs can guide the design of multi-layer transportation solutions, from optimizing elevator operations at peak times to integrating 3D route-finding algorithms into personal navigation devices.

- Public Health and Sustainability: Quantifying vertical activity has implications for health monitoring (energy expenditure and activity levels) and for estimating the environmental impact of vertical transportation (e.g., elevator energy consumption). A more accurate assessment of these factors can aid policymakers in developing interventions that enhance well-being and reduce carbon footprints.

8.3 Strengths and Limitations:

Strengths: A key strength of this research lies in the scale, diversity, and real-world complexity of the data analyzed. The incorporation of multiple sensor modalities (barometers, accelerometers, Wi-Fi signals) and the application of advanced machine learning methods have enabled robust, accurate,

context-rich understanding. The integration of topological analysis and network theory provides a novel perspective that links structural aspects of urban environments to observed human behavior.

Limitations: Nevertheless, certain limitations exist. The demographics focusing on younger participants (students) may limit the generalizability of observed patterns to other age groups or socioeconomic strata. The absence of longer-term or seasonally varied data may not capture temporal shifts in vertical mobility preferences. Variability in sensor accuracy, resolution, and placement may have also constrained the analysis. Moreover, the COVID-19 pandemic context during some data collection periods could have influenced typical movement patterns. Future studies should seek broader population samples, integrate richer environmental metadata, and consider longitudinal designs to address these constraints.

8.4 Future Research Directions

Given the nascent nature of vertical mobility studies, numerous avenues for future research emerge:

- Longitudinal Studies: Examine how vertical mobility patterns evolve over months or years, capturing changes due to infrastructure upgrades, policy interventions, or shifting socioeconomic conditions.
- Diverse Environments and Populations: Expand analyses beyond a single city-state to include diverse urban contexts, varied climatic conditions, and more heterogeneous populations.
- Advanced Modeling Techniques: Utilize computational models, agent-based simulations, or machine learning-driven predictions to forecast the effects of new vertical corridors, integrated developments, or emerging vertical transit systems.
- Socio-Spatial Networks: Integrate socio-demographic data, user preferences, and social networks to understand how human interactions and behavioral factors intersect with built environment design, further refining guidelines for inclusive, resilient urban growth.

8.5 Conclusion

In summary, this thesis demonstrated that vertical mobility is a crucial, quantifiable, and interpretable aspect of human movement in dense, three-

dimensional urban environments. By systematically examining the factors influencing barometric sensing, developing a robust analytical framework for vertical displacement identification, and characterizing both city-scale and building-scale vertical mobility, we have provided new insights into the complexity of human movement patterns. These findings not only advance our theoretical understanding of mobility but also offer practical guidance for designing more connected, health-promoting, and sustainable cities. As urban landscapes continue to grow vertically, our work lays the groundwork for future studies that fully embrace the volumetric nature of human habitats and the behaviors they inspire.

Bibliography

[1] E. Wilhelm, S. Siby, Y. Zhou, X. J. S. Ashok, M. Jayasuriya, S. Foong, J. Kee, K. L. Wood, and N. O. Tippenhauer, "Wearable environmental sensors and infrastructure for mobile large-scale urban deployment," *IEEE Sensors Journal*, vol. 16, no. 22, pp. 8111–8123, 2016.

[2] Y. Zhou, J. Wang, P. Shi, D. Dahlmeier, N. Tippenhauer, and E. Wilhelm, "Power-saving transportation mode identification for large-scale applications," 2017.

[3] M. Straczkiewicz and J.-P. Onnela, "A systematic review of human activity recognition using smartphones," *arXiv preprint arXiv:1910.03970*, 2019.

[4] United Nations, "The world's cities in 2018—data booklet." `https://digitallibrary.un.org/record/3799524`, 12 2018. (accessed Feb. 22, 2022).

[5] W. Bank, "Urban development," 2022. (accessed Oct. 30, 2022).

[6] L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences*, vol. 104, no. 17, pp. 7301–7306, 2007.

[7] T. Schröpfer, *Dense + Green Cities*. Birkhäuser, Dec. 2019.

[8] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, "Human mobility: Models and applications," *Physics Reports*, vol. 734, pp. 1–74, 2018.

[9] A. Manivannan, W. C. B. Chin, A. Barrat, and R. Bouffanais, "On the challenges and potential of using barometric sensors to track human activity," *Sensors*, vol. 20, no. 23, p. 6786, 2020.

[10] A. Manivannan, E. J. Willemse, B. Balamurali, W. C. B. Chin, Y. Zhou, B. Tunçer, A. Barrat, and R. Bouffanais, "A framework for the identification of human vertical displacement activity based on multi-sensor data," *IEEE Sensors Journal*, vol. 22, no. 8, pp. 8011–8029, 2022.

[11] A. Manivannan, W. C. B. Chin, S. Gopalakrishnan, D. K. Wong, T. Schroepfer, and R. Bouffanais, "Effects of the interplay between topology and function of an integrated urban development on patterns of user movement," *Scientific Reports*, vol. 14, no. 1, p. 7021, 2024.

[12] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[13] M. B. Del Rosario, S. J. Redmond, and N. H. Lovell, "Tracking the evolution of smartphone sensing for monitoring human movement," *Sensors*, vol. 15, no. 8, pp. 18901–18933, 2015.

[14] C. Bollmeyer, T. Esemann, H. Gehring, and H. Hellbrück, "Precise indoor altitude estimation based on differential barometric sensing for wireless medical applications," in *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–6, IEEE, 2013.

[15] S. Vanini, F. Faraci, A. Ferrari, and S. Giordano, "Using barometric pressure data to recognize vertical displacement activities on smartphones," *Computer Communications*, vol. 87, pp. 37–48, 2016.

[16] N. Pannurat, S. Thiemjarus, E. Nantajeewarawat, and I. Anantavrasilp, "Analysis of optimal sensor positions for activity classification and application on a different data collection scenario," *Sensors (Switzerland)*, vol. 17, no. 4, 2017.

[17] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pp. 4–pp, IEEE, 2006.

[18] M. Liu, H. Li, Y. Wang, F. Li, and X. Chen, "Double-windows-based motion recognition in multi-floor buildings assisted by a built-in barometer," *Sensors*, vol. 18, no. 4, p. 1061, 2018.

[19] Z. Xu, J. Wei, J. Zhu, and W. Yang, "A robust floor localization method using inertial and barometer measurements," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, IEEE, 2017.

[20] C. Wang, W. Lu, M. R. Narayanan, D. C. W. Chang, S. R. Lord, S. J. Redmond, and N. H. Lovell, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2302–2311, 2016.

[21] J. Hendiger, M. Chludzińska, and P. Zietek, "Influence of the pressure difference and door swing on heavy contaminants migration between rooms," *PloS one*, vol. 11, no. 5, p. e0155159, 2016.

[22] B.-J. Ho, P. Martin, P. Swaminathan, and M. Srivastava, "From pressure to path: Barometer-based vehicle tracking," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pp. 65–74, ACM, 2015.

[23] K. Muralidharan, A. J. Khan, A. Misra, R. K. Balan, and S. Agarwal, "Barometric phone sensors: More hype than hope!," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, p. 12, ACM, 2014.

[24] M. B. Del Rosario, K. Wang, J. Wang, Y. Liu, M. Brodie, K. Delbaere, N. H. Lovell, S. R. Lord, and S. J. Redmond, "A comparison of activity classification in younger and older cohorts using a smartphone," *Physiological measurement*, vol. 35, no. 11, p. 2269, 2014.

[25] X. Bao, Z. Xiong, S. Sheng, Y. Dai, S. Bao, and J. Liu, "Barometer measurement error modeling and correction for uah altitude tracking," in *2017 29th Chinese Control And Decision Conference (CCDC)*, pp. 3166–3171, IEEE, 2017.

[26] K. Sankaran, M. Zhu, X. F. Guo, A. L. Ananda, M. C. Chan, and L.-S. Peh, "Using mobile phone barometer for low-power transporta-

tion context detection," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, pp. 191–205, ACM, 2014.

[27] M. Tachikawa, T. Maekawa, and Y. Matsushita, "Predicting location semantics combining active and passive sensing with environment-independent classifier," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 220–231, ACM, 2016.

[28] G. Pipelidis, O. R. M. Rad, D. Iwaszczuk, C. Prehofer, and U. Hugentobler, "A novel approach for dynamic vertical indoor mapping through crowd-sourced smartphone sensor data," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, IEEE, 2017.

[29] C. Cenedese and P. J. Smith, "Atmospheric pressure and wind." Available online: `https://www.britannica.com/science/climate-meteorology/Atmospheric-pressure-and-wind`, 5 2019. (accessed on 20/01/2020).

[30] D. F. Young, B. R. Munson, T. H. Okiishi, and W. W. Huebsch, *A brief introduction to fluid mechanics*. John Wiley & Sons, 2010.

[31] H. Ye, T. Gu, X. Tao, and J. Lu, "Scalable floor localization using barometer on smartphone," *Wireless Communications and Mobile Computing*, vol. 16, no. 16, pp. 2557–2571, 2016.

[32] F. Zhao, H. Luo, X. Zhao, Z. Pang, and H. Park, "Hyfi: Hybrid floor identification based on wireless fingerprinting and barometric pressure," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 330–341, 2015.

[33] Trane, "Commercial Building Pressurization," *TRANE engineers newsletter*, vol. 31, no. 2, 2002.

[34] M. Wu, P. H. Pathak, and P. Mohapatra, "Monitoring building door events using barometer sensor in smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 319–323, ACM, 2015.

[35] N. Zhang and D. Zhou, "Numerical analysis of the pressure variation in subway tunnel when an on-fire train runs at different speed," in *The 2016 World Congress on Advances in Civil, Environmental, and Materials Research (ACEM16)*, pp. 1–6, IASEM, 2016.

[36] P. Reinke, M. Flueckiger, and T. Wicht, "Aerodynamics and ventilation in rail tunnels may 2015," 5 2015.

[37] "Atmospheric pressure." Available online: `https://www.britannica.com/science/atmospheric-pressure`, 2 2019. (accessed on 20/11/2019).

[38] H. Xia, X. Wang, Y. Qiao, J. Jian, and Y. Chang, "Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning," *Sensors*, vol. 15, no. 4, pp. 7857–7877, 2015.

[39] F. Haque, V. Dehghanian, A. O. Fapojuwo, and J. Nielsen, "A sensor fusion-based framework for floor localization," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 623–631, 2018.

[40] K. Sagawa, T. Ishihara, A. Ina, and H. Inooka, "Classification of human moving patterns using air pressure and acceleration," in *IECON'98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No. 98CH36200)*, vol. 2, pp. 1214–1219, IEEE, 1998.

[41] Y. Chen and C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[42] J. Mäntyjärvi, J. Himberg, and T. Seppänen, "Recognizing human motion with multiple acceleration sensors," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 747–752, 2001.

[43] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," *Sensors (Switzerland)*, vol. 19, no. 7, 2019.

[44] K. Kunze and P. Lukowicz, "Sensor placement variations in wearable activity recognition," *IEEE Pervasive Computing*, vol. 13, no. 4, pp. 32–41, 2014.

[45] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 386–403, 2017.

[46] L. Cong, J. Tian, and H. Qin, "A practical floor localization algorithm based on multifeature motion mode recognition utilizing fm radio signals and inertial sensors," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8806–8819, 2020.

[47] S. Boim, G. Even-Tzur, and I. Klein, "Height difference determination using smartphones based accelerometers," *IEEE Sensors Journal*, 2021.

[48] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, and K. Aberer, "Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach," *Proceedings - International Symposium on Wearable Computers, ISWC*, pp. 17–24, 2012.

[49] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data BT - UbiComp 2002: Ubiquitous Computing," *UbiComp 2002: Ubiquitous Computing*, vol. 3001, no. Chapter 1, pp. 1–17, 2004.

[50] A. Moncada-Torres, K. Leuenberger, R. Gonzenbach, A. Luft, and R. Gassert, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," *Physiological measurement*, vol. 35, no. 7, p. 1245, 2014.

[51] A. El Halabi and H. Artail, "Integrating pressure and accelerometer sensing for improved activity recognition on smartphones," in *2013 Third International Conference on Communications and Information Technology (ICCIT)*, pp. 121–125, IEEE, 2013.

[52] M. Ji, J. Liu, X. Xu, and Z. Lu, "The improved 3d pedestrian positioning system based on foot-mounted inertial sensor," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25051–25060, 2020.

[53] Y. Zhao, J. Liang, Y. Cui, X. Sha, and W. J. Li, "Adaptive 3d position estimation of pedestrians by wearing one ankle sensor," *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11642–11651, 2020.

[54] M. Elhoushi, J. Georgy, A. Wahdan, M. Korenberg, and A. Noureldin, "Using portable device sensors to recognize height changing modes of motion," in *2014 Ieee International Instrumentation and Measurement Technology Conference (I2mtc) Proceedings*, pp. 477–481, IEEE, 2014.

[55] M. Monteiro and A. C. Martí, "Using smartphone pressure sensors to measure vertical velocities of elevators, stairways, and drones," *arXiv preprint arXiv:1607.00363*, 2016.

[56] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3 – 11, 2019. Deep Learning for Pattern Recognition.

[57] C. Lang and S. Kaiser, "Classifying elevators and escalators in 3d pedestrian indoor navigation using foot-mounted sensors," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–7, IEEE, 2018.

[58] M. U. Kraemer, A. Sadilek, Q. Zhang, N. A. Marchal, G. Tuli, E. L. Cohn, Y. Hswen, T. A. Perkins, D. L. Smith, R. C. Reiner Jr, *et al.*, "Mapping global variation in human mobility," *Nature Human Behaviour*, vol. 4, no. 8, pp. 800–810, 2020.

[59] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, 2017.

[60] S. Gopalakrishnan, D. Wong, A. Manivannan, R. Bouffanais, and T. Schroepfer, "Mapping emergent patterns of movement and space use in vertically integrated urban developments," in *12th Annual Symposium on Simulation for Architecture and Urban Design (SimAUD)*, pp. 1–8, 2021.

[61] X. Yang, Z. Zhao, and S. Lu, "Exploring spatial-temporal patterns of urban human mobility hotspots," *Sustainability*, vol. 8, no. 7, p. 674, 2016.

[62] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 186–194, 2012.

[63] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 157–165, 2011.

[64] K. K. Jahromi, M. Zignani, S. Gaito, and G. P. Rossi, "Simulating human mobility patterns in urban areas," *Simulation Modelling Practice and Theory*, vol. 62, pp. 137–156, 2016.

[65] D. E. Boyce and H. C. Williams, *Forecasting urban travel: Past, present and future*. Edward Elgar Publishing, 2015.

[66] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González, "Understanding road usage patterns in urban areas," *Scientific reports*, vol. 2, no. 1, p. 1001, 2012.

[67] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, "On the use of human mobility proxies for modeling epidemics," *PLoS computational biology*, vol. 10, no. 7, p. e1003716, 2014.

[68] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, "Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions," *PLoS medicine*, vol. 4, no. 1, p. e13, 2007.

[69] A. Apolloni, C. Poletto, J. J. Ramasco, P. Jensen, and V. Colizza, "Metapopulation epidemic models with heterogeneous mixing and travel behaviour," *Theoretical Biology and Medical Modelling*, vol. 11, pp. 1–26, 2014.

[70] K. Zhao, S. Tarkoma, S. Liu, and H. Vo, "Urban human mobility data mining: An overview," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 1911–1920, IEEE, 2016.

[71] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th international conference on World wide web*, pp. 791–800, 2009.

[72] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," in *2010 IEEE international conference on data mining*, pp. 971–976, IEEE, 2010.

[73] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[74] M. Lenormand, T. Louail, O. G. Cantú-Ros, M. Picornell, R. Herranz, J. M. Arias, M. Barthelemy, M. S. Miguel, and J. J. Ramasco, "Influence of sociodemographic characteristics on human mobility," *Scientific reports*, vol. 5, no. 1, p. 10075, 2015.

[75] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[76] J. P. Bagrow and Y.-R. Lin, "Mesoscopic structure and social aspects of human mobility," *PloS one*, vol. 7, no. 5, p. e37676, 2012.

[77] L. Alessandretti, P. Sapiezynski, S. Lehmann, and A. Baronchelli, "Multi-scale spatio-temporal analysis of human mobility," *PLoS ONE*, vol. 12, p. e0171686, feb 2017.

[78] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási, "Returners and explorers dichotomy in human mobility," *Nature communications*, vol. 6, no. 1, p. 8166, 2015.

[79] S. M. Iacus, C. Santamaria, F. Sermi, S. Spyratos, D. Tarchi, and M. Vespe, "Human mobility and covid-19 initial dynamics," *Nonlinear Dynamics*, vol. 101, pp. 1901–1919, 2020.

[80] L. Pappalardo and F. Simini, "Data-driven generation of spatio-temporal routines in human mobility," *Data Mining and Knowledge Discovery*, vol. 32, no. 3, pp. 787–829, 2018.

[81] T. Liu, Z. Yang, Y. Zhao, C. Wu, Z. Zhou, and Y. Liu, "Temporal understanding of human mobility: A multi-time scale analysis," *PloS one*, vol. 13, no. 11, p. e0207697, 2018.

[82] F. Xu, Y. Li, D. Jin, J. Lu, and C. Song, "Emergence of urban growth patterns from human mobility behavior," *Nature Computational Science*, vol. 1, no. 12, pp. 791–800, 2021.

[83] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature physics*, vol. 6, no. 10, pp. 818–823, 2010.

[84] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *Physical Review E*, vol. 80, no. 2, p. 021136, 2009.

[85] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger, "Human mobility modeling at metropolitan scales," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 239–252, 2012.

[86] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *Journal of Statistical Physics*, vol. 151, no. 1, pp. 304–318, 2013.

[87] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Scientific reports*, vol. 3, no. 1, p. 2923, 2013.

[88] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[89] A. Cuttone, S. Lehmann, and M. C. González, "Understanding predictability and exploration in human mobility," *EPJ Data Science*, vol. 7, pp. 1–17, 2018.

[90] M. G. Meekan, C. M. Duarte, J. Fernández-Gracia, M. Thums, A. M. Sequeira, R. Harcourt, and V. M. Eguíluz, "The ecology of human mobility," *Trends in ecology & evolution*, vol. 32, no. 3, pp. 198–210, 2017.

[91] C. Marchetti, "Anthropological invariants in travel behavior," *Technological forecasting and social change*, vol. 47, no. 1, pp. 75–88, 1994.

[92] R. Kölbl and D. Helbing, "Energy laws in human travel behaviour," *New Journal of Physics*, vol. 5, no. 1, p. 48, 2003.

[93] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, and A. Baronchelli, "Evidence for a conserved quantity in human mobility," *Nature human behaviour*, vol. 2, no. 7, pp. 485–491, 2018.

[94] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PloS one*, vol. 7, no. 5, p. e37027, 2012.

[95] M. Schläpfer, L. Dong, K. O'Keeffe, P. Santi, M. Szell, H. Salat, S. Anklesaria, M. Vazifeh, C. Ratti, and G. B. West, "The universal visitation law of human mobility," *Nature*, vol. 593, no. 7860, pp. 522–527, 2021.

[96] X. Liang, J. Zhao, L. Dong, and K. Xu, "Unraveling the origin of exponential law in intra-urban human mobility," *Scientific reports*, vol. 3, no. 1, p. 2983, 2013.

[97] W. Wang, L. Pan, N. Yuan, S. Zhang, and D. Liu, "A comparative analysis of intra-city human mobility by taxi," *Physica A: Statistical Mechanics and its Applications*, vol. 420, pp. 134–147, 2015.

[98] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, no. 1, pp. 1–5, 2013.

[99] S. D. of Statistics, *Census of Population, 2020: Statistical Release 2 : Households, Geographic Distribution, Transport, and Difficulty in Basic Activities*. Department of Statistics, Ministry of Trade and Insdustry, Republic of Singapore, 2021.

[100] C. Zhong, S. M. Arisona, X. Huang, and G. Schmitt, "Identifying spatial structure of urban functional centers using travel survey data: A case study of singapore," 2013.

[101] C. Zhong, M. Batty, E. Manley, J. Wang, Z. Wang, F. Chen, and G. Schmitt, "Variability in regularity: Mining temporal mobility patterns in london, singapore and beijing using smart-card data," *PloS one*, vol. 11, no. 2, p. e0149222, 2016.

[102] H. Poonawala, V. Kolar, S. Blandin, L. Wynter, and S. Sahu, "Singapore in motion: Insights on public transport service level through farecard and mobile data analytics," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and data mining*, pp. 589–598, 2016.

[103] Y. Xu, A. Belyi, I. Bojic, and C. Ratti, "Human mobility and socioeconomic status: Analysis of singapore and boston," *Computers, Environment and Urban Systems*, vol. 72, pp. 51–67, 2018.

[104] W. C. B. Chin, C.-C. Feng, C.-H. Leong, J. Pang, H. E. Clapham, A. Nara, M.-H. Tsou, and Y.-C. Wang, "Integrating local and neighboring area influences into vulnerability modeling of infectious diseases in singapore," *International Journal of Applied Earth Observation and Geoinformation*, vol. 121, p. 103376, 2023.

[105] Q. Chen, I.-T. Chuang, and A. Poorthuis, "Entangled footprints: Understanding urban neighbourhoods by measuring distance, diversity, and direction of flows in singapore," *Computers, Environment and Urban Systems*, vol. 90, p. 101708, 2021.

[106] J. H. Kurniawan, C. Ong, and L. Cheah, "Examining values and influences affecting public expectations of future urban mobility: A singapore case study," *Transport Policy*, vol. 66, pp. 66–75, 2018.

[107] A. J. Hakim, K. R. Victory, J. R. Chevinsky, M. A. Hast, D. Weikum, L. Kazazian, S. Mirza, R. Bhatkoti, M. Schmitz, M. Lynch, *et al.*, "Mitigation policies, community mobility, and covid-19 case counts in australia, japan, hong kong, and singapore," *Public Health*, vol. 194, pp. 238–244, 2021.

[108] J. Li, F. Tartarini, *et al.*, "Changes in air quality during the covid-19 lockdown in singapore and associations with human mobility trends," *Aerosol and Air Quality Research*, vol. 20, no. 8, pp. 1748–1758, 2020.

[109] S. E. Bibri, J. Krogstie, and M. Kärrholm, "Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability," *Developments in the Built Environment*, vol. 4, p. 100021, 2020.

[110] R. Freestone, "Better planning and research for mixed-use developments," *Australian Planner*, vol. 45, no. 1, pp. 14–15, 2008.

[111] I. Mateo-Babiano and S. Darchen, "Vertical mixed use communities, a compact city model?," in *6th State of Australian Cities Conference*, November 2013.

[112] J. Rabianski, K. Gibler, O. A. Tidwell, and J. S. Clements III, "Mixed-use development: A call for research," *Journal of Real Estate Literature*, vol. 17, no. 2, pp. 205–230, 2009.

[113] Y. Liao, J. Gil, R. H. Pereira, S. Yeh, and V. Verendel, "Disparities in travel times between car and transit: Spatiotemporal patterns in cities," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[114] M. Jemmali, L. K. B. Melhim, M. T. Alharbi, A. Bajahzar, and M. N. Omri, "Smart-parking management algorithms in smart city," *Scientific Reports*, vol. 12, no. 1, pp. 1–15, 2022.

[115] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, "Smart cities of the future," *The European Physical Journal Special Topics*, vol. 214, no. 1, pp. 481–518, 2012.

[116] J. Al Dakheel, C. Del Pero, N. Aste, and F. Leonforte, "Smart buildings features and key performance indicators: A review," *Sustainable Cities and Society*, vol. 61, p. 102328, 2020.

[117] J. Nembrini and D. Lalanne, "Human-building interaction: When the machine becomes a building," in *IFIP Conference on Human-Computer Interaction - INTERACT 2017*, pp. 348–369, Springer, 2017.

[118] M. Simpson, K.-F. Richter, J. O. Wallgrün, and A. Klippel, "Quantifying space, understanding minds: A visual summary approach," *Journal of Spatial Information Science*, no. 14, pp. 95–136, 2017.

[119] A. Natapov, S. Kuliga, R. C. Dalton, and C. Hölscher, "Linking building-circulation typology and wayfinding: Design, spatial analysis, and anticipated wayfinding difficulty of circulation types," *Architectural Science Review*, vol. 63, no. 1, pp. 34–46, 2020.

[120] M. Rahimi, M. R. Malek, C. Claramunt, and T. Le Pors, "A topology-based graph data model for indoor spatial-social networking," *International Journal of Geographical Information Science*, vol. 35, no. 12, pp. 2517–2539, 2021.

[121] V. M. Netto, "'What is space syntax not?' Reflections on space syntax as sociospatial theory," *Urban Design International*, vol. 21, no. 1, pp. 25–40, 2016.

[122] K. Karimi, "A configurational approach to analytical urban design: 'Space syntax' methodology," *Urban Design International*, vol. 17, no. 4, pp. 297–318, 2012.

[123] S. K. Ravulaparthy and K. G. Goulias, "Characterizing the composition of economic activities in central locations: Graph-theoretic approach to urban network analysis," *Transportation Research Record*, vol. 2430, no. 1, pp. 95–104, 2014.

[124] S. Porta, P. Crucitti, and V. Latora, "The network analysis of urban streets: A primal approach," *Environment and Planning B: Planning and Design*, vol. 33, no. 5, pp. 705–725, 2006.

[125] M. Batty, *The New Science of Cities*. The MIT press, 2013.

[126] A.-L. Barabási and M. Pósfai, *Network Science*. Cambridge: Cambridge University Press, 2016.

[127] A. D. S. Srikanth, W. C. B. Chin, R. Bouffanais, and T. Schroepfer, "Complexity science for urban solutions," in *Artificial Intelligence in Urban Planning and Design: Technologies, Implementation, and Impacts* (I. As, P. Basu, and P. Talwar, eds.), ch. 3, pp. 39–58, Elsevier, 5 2022.

[128] D. Chang, "Spatial choice and preference in multilevel movement networks," *Environment and Behavior*, vol. 34, no. 5, pp. 582–615, 2002.

[129] B. Tunçer and F. Benita, "Data-driven thinking for measuring the human experience in the built environment," *International Journal of Architectural Computing*, p. 14780771211025142, 2021.

[130] A. Mashhadi, U. G. Acer, A. Boran, P. M. Scholl, C. Forlivesi, G. Vanderhulst, and F. Kawsar, "Exploring space syntax on entrepreneurial opportunities with Wi-Fi analytics," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 658–669, 2016.

[131] A. Willis, N. Gjersoe, C. Havard, J. Kerridge, and R. Kukla, "Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments," *Environment and Planning B: Planning and Design*, vol. 31, no. 6, pp. 805–828, 2004.

[132] M. Serra and B. Hillier, "Angular and metric distance in road network analysis: A nationwide correlation study," *Computers, Environment and Urban Systems*, vol. 74, pp. 194–207, 2019.

[133] L. Zhang, A. Chiaradia, and Y. Zhuang, "A configurational accessibility study of road and metro network in shanghai, china," in *Recent Developments in Chinese Urban Planning: Selected Papers from the 8th International Association for China Planning Conference, Guangzhou, China, June 21-22, 2014*, pp. 219–245, Springer, 2015.

[134] C. H. Cooper, "Spatial localization of closeness and betweenness measures: a self-contradictory but useful form of network analysis," *International Journal of Geographical Information Science*, vol. 29, no. 8, pp. 1293–1309, 2015.

[135] L. Zhang and A. Chiaradia, "From axial to pedestrian path-centre line: The case of 3d pedestrian network in hong kong, central," in *The 12th International Space Syntax Symposium (12SSS)*, Beijing JiaoTong University., 2019.

[136] "Home," 10 2017.

[137] Skyhook, "Skyhook location services," 2021. (accessed Feb. 22, 2022).

[138] A. Graser, "MovingPandas: Efficient Structures for Movement Data in Python," *GI_Forum – Journal of Geographic Information Science*, vol. 7, no. 1, pp. 54–68, 2019.

[139] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.