12-2023

Characterization of Human Mobility from Cellular Data

Zaid Matloub
Florida Institute of Technology, zmatloub2010@my.fit.edu

Characterization of Human Mobility from Cellular Data

By

Zaid Matloub

B.S. in Electronics and Communications Engineering, University of Baghdad, 2009
M.S. in Electrical Engineering, Florida Institute of Technology, 2012

A dissertation submitted to the College of Engineering and Science of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Electrical Engineering

Melbourne, Florida
December, 2023

We the undersigned committee hereby approve the attached dissertation,

CHARACTERIZATION OF HUMAN MOBILITY FROM CELLULAR DATA

By

Zaid Matloub

_____

Ivica Kostanic, Ph.D.
Associate Professor
Electrical Engineering and Computer Science
Major Advisor

_____

Carlos E. Otero, Ph.D.
Professor
Electrical Engineering and Computer Science

_____

Josko Zec, Ph.D.
Graduate Faculty
Electrical Engineering and Computer Science

_____

Ersoy Subasi, Ph.D.
Associate Professor
College of Aeronautics

_____

Brian Lail, Ph.D.
Professor and Department Head
Electrical Engineering and Computer Science

Abstract

CHARACTERIZATION OF HUMAN MOBILITY FROM CELLULAR DATA

By

Zaid Matloub

Dissertation Advisor: Ivica Kostanic, Ph.D.

This dissertation investigates human mobility patterns using crowd-sourced cellular network data from different Metropolitan Statistical Areas (MSAs) in the United States, spanning the Houston, New York-Newark, NJ City, and 13 other significant MSAs. By focusing on prominent spatial mobility parameters highlighted in existing literature, the study unveils consistent findings regarding the predictability of human mobility across diverse time scales and geographic regions. The research underscores the significance of selecting appropriate sampling thresholds based on the mobility parameters being examined, the size of the dataset, and available computational resources. Through a meticulous analysis, it emerges that while values such as mean and standard deviation may fluctuate based on sampling thresholds, the distribution patterns of mobility parameters remain notably consistent. Diving deeper, the dissertation classifies MSAs into two primary groups based on observed travel patterns: inland and coastal MSAs, revealing distinct weekly travel trends for each group. These comprehensive insights not only contribute to a foundational understanding of human mobility across MSAs but also highlight the potential for influencing urban planning and business decisions when combined with supplementary data sources.

Table of Contents

List of Figures

List of Tables

Acknowledgement

My deep appreciation and sincere gratitude go to my advisor, Dr. Ivica Kostanic, for his consistent guidance, encouragement, and invaluable advice. His expertise and perspective on life not only enhance my academic experience but also profoundly influence my personal growth. I am genuinely honored to call him my mentor.

I extend my heartfelt thanks to my parents, who shower me with unconditional love and support. Their teachings shape my discipline and ambition, and they continually make sacrifices that enable me to reach such significant milestones.

To my sister, your unwavering support and consistent encouragement are my anchors. You cheer me up during tough times and the moments of laughter we share over calls bring much-needed relief. Your gestures, both big and small, play a vital role in this journey.

I would be remiss if I didn't acknowledge my dear friends, Dr. Ismail Alkhouri and Rand Alattar. Their friendship and consistent support have been instrumental in my academic endeavors, making them an integral part of this milestone.

Special appreciation is reserved for Better world Analytics USA. Their invaluable contribution of the data, as well as the processing tools and resources, were key components that facilitated the successful completion of this research.

To all of you, thank you for standing by me and for being integral to this important chapter of my life.

Dedication

To Samir and Eman, my beloved parents and pillars of strength, and to Maryam, my sister and confidant, this achievement is as much yours as it is mine.

.

Chapter 1
Introduction

In modern society, technological advancements have profoundly impacted individuals' daily lives, reshaping urban dynamics and lifestyles [1] [2]. While human mobility, encompassing activities such as commuting, shopping, and social engagements, might appear random at first, an analysis of mobility data uncovers inherent patterns. This understanding underscores the recurring nature of human mobility, driven by personal choices and societal ties [3] [4].

Historically, the study of human movement predominantly used tools like surveys. However, these approaches faced limitations due to their limited scope, cost, and time-consuming nature [5][6]. The technological era, marked by the widespread use of mobile devices, has introduced a transformative method of data collection [1][2]. In urban areas, these devices diligently track individuals' daily movements [7]. Among various location-based datasets, Call Detail Records (CDR) have become a key resource to study recent human mobility trends. CDRs, which include user IDs, cell tower locations, and timestamps, offer crucial insights into the spatial and temporal patterns of urban human activity [7]. Moreover, tech giants such as Google, Facebook, Apple, and other data providers leverage GPS data, granting a more detailed perspective on human movement, especially given GPS's higher spatial precision compared to CDRs.

Nevertheless, while consistent patterns can be discerned from short-term mobility data, unpredictable incidents like natural disasters can introduce abrupt shifts [3]. For instance, an earthquake might necessitate city evacuations, markedly changing movement trajectories [3]. In these scenarios, the ubiquity of mobile devices proves

essential for monitoring individual routes, thereby highlighting any irregularities in movement.

This comprehensive knowledge of human mobility, augmented by mobile data, plays a pivotal role across various fields. Areas such as urban planning, traffic forecasting, real-time event detection, disaster response, disease prevention, environmental assessments, and service provisioning by both governmental and private sectors are closely connected to our comprehension of mobility patterns [8][9]. The growing prevalence of mobile-broadband networks worldwide amplifies this relationship, with many regions seeing over 90% of their inhabitants accessing such networks [6]. Hence, the widespread use of mobile phones has become fundamental for investigating and addressing the nuances of human mobility in today's digital world, especially in developing nations.

Problem Statement

Modern cities are continuously evolving due to factors such as rapid urbanization, migratory movements, climate change, pandemics like COVID-19, the rise of social networks, and the increasing shift to remote work. In light of these changes, it is essential to delve into the intricate dynamics of urban spaces through the lens of human mobility [1], [7]. As highlighted earlier, traditional methods of collecting human mobility data have their shortcomings. They can be expensive, labor-intensive, and might not fully capture the vibrant and changing nature of growing cities [7]. Digital technologies, especially data from cellular users, offer a potential solution. However, a significant portion of the current literature is specialized in its approach and may not comprehensively represent the mobility patterns of a city's residents [4], [10].

Many mobility studies, on the other hand, tend to be regionally specific, overlooking the variety of urban populations and their unique movement behaviors. This specificity reveals a gap in the existing research. Often, these studies do not provide a complete perspective on how people traverse cities and manage their daily routines. Although a plethora of research utilizes cellular data to demonstrate urban mobility patterns, the diversity in methodologies makes it problematic to draw consistent conclusions across various cities. This underscores a pivotal research challenge: a uniform, adaptable, and all-encompassing approach to measure mobility in different urban environments.

Using crowd-sourced cellular data provides a deeper understanding of human mobility trends. Recognizing these patterns is crucial for effective urban management, touching on areas such as public safety, tourism, infrastructure, and transportation. For example, mapping daily movement routes can reveal connections between human mobility and its impact on the environment. Similarly, understanding transit patterns between vital hubs, like residences and workplaces, contributes significantly to urban planning and efficient traffic management. The primary objective of this research is to identify crucial mobility parameters and observe patterns across varied demographics, aiming to determine the consistency of these parameters across cities in the United States. Future studies may expand this investigation to an international scale.

Research Objectives

- *Mobility Parameters Identification and Validation*: Extract and validate a comprehensive set of spatial mobility parameters from existing literature, which includes but is not limited to Number of Visited Locations (N_LOC), Number of Unique Visited Locations (N_ULOC), Radius of Gyration (R_GYR), and Distance Traveled (D_TRAV).

- *Sampling Threshold Analysis*: Investigate the effects of varied spatial and time sampling thresholds on a set of studied mobility parameters, across daily, weekly, and monthly timeframes.
- *Geographical Expansion*: Expand the scope of research to incorporate diverse geographies, focusing especially on large Metropolitan Statistical Areas (MSAs) in the United States based on recent population estimates. This ensures the research encapsulates varied user profiles and is not restricted to smaller geographic pockets.
- *Comparative Analysis*: Analyze the consistency of mobility parameters within and across different MSAs in the U.S. This will help in noting patterns, similarities, and variations in terms of individual and group mobility."

Dissertation Outline

Chapter 1: Introduction: Introduces the research focus and objectives of this dissertation.

Chapter 2: Literature Review: Examines the existing body of literature relevant to the study.

Chapter 3: Dataset Description: Details the dataset used for mobility analysis, including its format and the detailed preprocessing steps involved.

Chapter 4: Mobility Parameters: Definition and Analyses Results: Analyzes key parameters related to human mobility, highlighting the spatial aspects. Each parameter's definition is articulated, with comprehensive assessments spanning three observation periods: daily, weekly, and monthly.

Chapter 5: Comparative Study in 15 Top MSAs: Investigates mobility patterns across the 15 largest Metropolitan Statistical Areas (MSAs) in the U.S., employing specific spatial and derived mobility parameters. The chapter offers a detailed comparative analysis, grouping MSAs by analogous mobility patterns.

Chapter 6: Conclusion and Future Work: Summarizes the findings and discusses potential avenues for extending the research presented in the dissertation.

Chapter 2
Literature Review

In the review of the literature on human mobility studies using cellular data, studies are identified and classified into nine distinct categories: General Models for Human Mobility, Urban Planning, Points of Interest (POIs), Origin-Destination Matrix (OD Matrix), User Profiling, Group Movement Patterns and Epidemic Spread. Within each category, the analysis details how the selected studies align with the respective research topic. Additionally, the limitations of these studies are presented.

General Models for Human Mobility

In the realm of human mobility models, a wealth of studies have emerged from the literature, showcasing theoretical frameworks that leverage statistical techniques to decipher user trajectories. These investigations accentuate the invaluable role of cellular data, particularly when exploring mobility patterns in major cities across the globe. A pivotal facet in these studies is the access to comprehensive datasets, which mobile operators can supply. This access not only augments the depth and breadth of mobility studies but also paves the way for advancements in areas like intelligent transport systems and the evolution of Smart Cities.

Kang et al. (2010) pioneered several spatiotemporal analysis and preprocessing techniques within the time geography framework to scrutinize individual and aggregated mobility patterns. By examining cell phone usage across varying days and times, they gleaned insights into mobility trends of millions of users. Focusing on a large Chinese city, their data segmentation, based on social factors like age and

gender, revealed consistent activity distances for individuals, irrespective of gender. Notably, middle-aged and younger individuals demonstrated more pronounced mobility compared to the elderly and adolescents [1].

In a parallel vein, Sevtsuk and Ratti (2010) delved into urban mobility patterns in Rome, Italy. Their study illuminated the remarkable consistency in urban mobility, regardless of the day or hour. Such findings offer invaluable insights for those at the nexus of urban planning and transportation, equipping them to better comprehend modern city dynamics [11].

Moving into empirical analysis, Song et al. (2010) collected mobile phone trace data to investigate continuous-time random walk (CTRW) models. By comparing these models to actual mobility patterns from two datasets, they proposed a nuanced model, underpinned by rules governing human movement. This model not only resonated with observed scaling laws but also facilitated analytical forecasting of significant scaling exponents [3].

In a visualization-centric approach, Pu et al. (2011) designed a system comprising three modules to depict population mobility patterns from a massive dataset collected in a major Chinese city. Their innovative Voronoi-diagram-based visual encoding method illuminated unique facets of mobile data, unearthing intriguing insights into human movement patterns [12].

Isaacman et al. (2012) introduced the WHERE model, a strategic tool delineating movement within metropolitan regions. Drawing upon call detail records, the model generates synthetic data for a fictitious population, marrying the need for authentic human mobility insights with privacy considerations. Their model was rigorously tested against extensive data from New York and Los Angeles, solidifying its efficacy [13].

Schneider et al. (2013) delved into the intricacies of daily mobility patterns, utilizing Markov chains to dissect the temporal and spatial trajectories of thousands as unique networks. Drawing from three diverse datasets – two from Paris (a survey and mobile phone billing data) and a Chicago-based survey – they identified 17 recurrent mobility networks. These networks, grounded on certain criteria, encapsulated up to 90% of individuals in both survey and mobile phone datasets spanning different countries [14].

Sun et al. (2016) shifted focus to user mobility, casting it under the lens of 4G data traffic. Leveraging a Hadoop-driven mobile big data platform and a robust mobility analysis framework, they parsed a dataset from a premier 4G cellular network in China. Their scrutiny not only highlighted the distinctions between findings from CDR/3G data and their 4G analyses but also posited that 4G data availed enhanced mobility and location granularity [15].

Zhao et al. (2017) too, harnessed 4G cellular network data to craft a user mobility model, addressing challenges spanning data collection, trajectory formation, noise elimination, and data storage. Their innovative model holds promise for areas such as urban planning, traffic forecasting, mobile computing, and resource optimization in radio networks [16].

Lind et al. (2017) introduced an innovative methodology to gauge human mobility. By extrapolating mobile subscribers' locations within a network's coverage, they formulated integrated mobility models founded on an augmented Kalman filter. The ingenuity of this method rests in estimating users' locations solely based on the network coverage cell they're linked to. When juxtaposed against GPS data and 271 diverse CDR records, the algorithm showcased impressive precision, deeming it a

potential tool for intelligent transport systems and location-based services. A key constraint, however, was its reliance solely on CDR data for location estimation [10].

Danafar, Piorkowski, and Krysczcuk (2017) underscored the digital footprints billions leave behind daily via mobile devices. These "Network Events (NEs)" are treasure troves of data, offering a window into prolonged human mobility patterns. While they present golden opportunities for mobile operators to roll out a spectrum of cost-effective services, from location-based offerings to traffic management, the coarse granularity and sparsity of NEs pose analytical challenges. Addressing this, the trio proposed a Bayesian strategy in [17], predicated on these network events, aiming to not only discern but also reconstruct mobility patterns, modes of transport, and frequent trajectories. Utilizing a dataset amassed from three test phones with thirty NE trace pairs (NE and corresponding GPS), their approach identified the most probable user route between initial and concluding NEs. As promising as their findings are, it beckons to be seen how this method fares with more expansive datasets [17].

Another notable human mobility prediction method for the users of mobile phones is presented by Hadachi et al. (2014). The method is based on an enhanced Markov Chain algorithm. As argued by the authors, it is challenging to predict the location of a user's mobility from the mobile phone data, which possesses a highly dynamic nature. According to the researchers in [18], applying the Markov Chain algorithm can efficiently manage these challenges. They combined two algorithms for predicting mobility: Local Prediction Algorithm (LPA) and Global Prediction Algorithm (GPA). They also opined that their proposed solution can be encouraging for the next generation of mobile networks and can also be used for optimizing the tracking systems and localization, road traffic, and the existing mobile network

infrastructure. However, considering unknown users, the accuracy of this method is not that high [18].

Recent studies have also focused on understanding human movement using crowd-sourced mobile phone data, proposing and testing theoretical and analytical models [9]. Knezevic et al. (2023) and Matloub et al. (2023) studied human movement in the Atlanta and Houston Metropolitan area using crowd-sourced network data, showing that human movement is highly predictable [19], and discussing how to select suitable sampling thresholds [20].

Urban Planning

Analyzing location data records can be instrumental for large cities around the world. These analyses allow for the projection of population growth, which can subsequently inform government decisions on city development, including land use, resolving mobility challenges, and other pertinent urban issues. There's a pressing need for longitudinal studies that delve deep into the dynamics of sprawling urban environments. These studies can offer invaluable comparisons between cities in developed and developing nations, highlighting differences in growth rates and socioeconomic contexts.

In their 2011 study, Becker et al. utilized call detail records to explore the flow of people in and out of Morristown, New Jersey (NJ), a suburban city in the U.S. with a population of approximately 20,000. The team extracted data from a prominent U.S. communications service provider, which consisted of 15 million voice and 26 million SMS records spanning 475,000 unique phones. This data was collected over a two-month period from November 29, 2009, to January 27, 2010. Through tabulation, statistical analysis, and visualization techniques, the researchers validated

the viability of their approach [21]. A comparison of these findings with movement patterns in Morristown, NJ derived from MDT data could be enlightening.

Simini et al.'s 2012 study introduced a stochastic model that, by only using population distribution data, could predict commuting and mobility flows. This "radiation model" demonstrated its ability to forecast a range of mobility patterns, from long-term migrations to regional communication volumes. Its independence from specific parameters means it's especially valuable in regions lacking prior mobility data, enhancing prediction accuracy [22].

Dash et al. in 2015 developed the Mobility Visualization System (MoVis) to depict human movement using call detail records. They analyzed three months' worth of mobile network data from 3.9 million users, totaling around 8 billion records. With this, they identified prominent locations people frequented. By extrapolating phone event trajectories between these locations, they could predict the start and end times of trips. They also mapped these journeys to existing transportation networks and visualized the most frequented routes between starting points and destinations. Their comprehensive visualization covered daily city mobility, distinguishing between weekdays and weekends. While their validation sample boasted high predictive accuracy for determining home and work locations [23], the study failed to clarify the criteria for participant selection and potential variability in outcomes.

Yang et al.'s 2020 paper emphasized the potential of frequently updated mobile phone signaling data in crafting smart and sustainable urban designs. They posited that call detail records could reveal variations in population density across city regions, assisting in the planning for urban growth and infrastructure distribution [24].

A significant point of discussion by Guo, Zhang, and Zhang in [25] was the critical nature of detecting urban traffic congestion. Traditional detection methodologies, such as manual surveys or fixed traffic data collection, often come with substantial costs and may not always be reliable. They championed the use of "Mobile Big Data" for this purpose, leveraging the fact that mobile phones in moving vehicles routinely connect to neighboring base stations, providing invaluable time and location data. This allows for more efficient and cost-effective traffic congestion detection. Notably, they proposed the use of HDFS (Hadoop Distributed File System) for data storage and Apache Spark for data processing. However, they left out details about whether the "mobile big data" used in their experiments was empirical or synthetic. The geographic location of the test data was also not mentioned [25].

Lastly, in a study by Xiang, Tu, and Huang [26], call detail records were used to examine urban dynamics in a major city in China. By tracking the movement of two million mobile phone users over a month, they discerned various calling habits and detected city 'barriers' like rivers or lakes. Such barriers might arise from reasons like undeveloped traffic systems, terrain challenges, or functional blockades. These barriers have significant implications for city connectivity and the efficiency of residents' travel. Recognizing these barriers is crucial for improving urban and transportation planning [26].

Points of Interests POIs: Home-Work locations

Understanding points of interests, such as homes or workplaces, is essential in discerning individuals' mobility patterns. However, to gain a comprehensive view of an individual's routine, considering all the places they frequent is crucial. This includes not only homes and workplaces but also venues of social gatherings, relaxation, and other activities. Differences in movement patterns during weekdays, weekends, and specific times can offer nuanced insights into societal behavior. Based

on this premise, four seminal studies have been highlighted, all striving to identify these primary locations. The studies, leveraging the concept that individuals are most likely found in specific locations, use various methods to map these primary spots, with clustering emerging as a predominant technique.

Csáji et al. (2013) embarked on an in-depth exploration of human behavior by utilizing an extensive dataset from Portugal. Their approach began by defining and calculating 50 features connected to calling habits. Following this, they conducted a correlation analysis and subsequently applied clustering and principal component analysis. Their objective was to pinpoint both home and work locations, emphasizing that individuals tend to concentrate their time at specific locations. Although the cellular data's exact nature remains unspecified, it's inferred that they employed CDR data, given their reliance on antenna locations for their analysis [4].

In 2014, Yang et al. presented a pioneering trajectory mining method that harnesses location gradients to discern between users' states of activity and rest. Drawing from an expansive dataset of 400 million anonymized records, representing around 10 million Beijing users, they introduced constructs like "HomeTime" and "WorkTime". These were derived from analyzing users' mobility patterns and were instrumental in recognizing significant locations. Their methodology, when applied to multi-day data, culminated in the identification of primary user locations, transcending traditional definitions of home or workplace. Their rigorous experiments corroborated the method's efficacy, indicating its superior performance when juxtaposed with existing methodologies [27].

Liu et al. (2017) introduced an algorithm tailored to identify mobile users' significant sites, drawing from cellular data that incorporates the Location Area Code (LAC) of serving base stations. Their findings illuminated a compelling correlation: the

similarity between two trajectories is intrinsically linked to their proximity within a location-based social network. The researchers deployed an unsupervised clustering approach, segmenting data into distinct social link categories. Their insights hold promise for myriad applications, especially in the spheres of urban planning and infrastructure development [28].

Highlighting the challenges inherent in using traditional metrics like SMS and call usage to determine primary locations, Tongsinoot et al. (2017) proposed an innovative methodology. Their approach amalgamated daily-aggregated internet usage data (G-CDR) with conventional CDRs to ascertain work locations. Simultaneously, to demarcate home locations, a sleep time analysis paradigm was introduced. Drawing from data spanning three months and encompassing over 4.3 million users in Bangkok, their findings underscored a pronounced urban trend: workplaces were predominantly centralized, while homes were more distributed [29].

Origin Destination Matrix (OD matrix)

The Origin-Destination (OD) matrix emerged as a significant area of interest from the literature review. This matrix quantifies the number of journeys from a starting point to a destination within a specific time and region. Frias-Martinez et al. (2012) emphasized that the OD matrix captures the movement of individuals between distinct geographical areas, effectively illustrating their paths [30]. Often, these geographical zones reveal insights about where individuals reside and operate.

In their study, the authors of [30] presented a novel technique for estimating commuting matrices. They harnessed data from two sources: the NSI mobility matrices for Madrid in 2009 and a dataset of cell phone call records (CDR) from the same period. By integrating optimization methods and a variation of Temporal

Association Rules, their methodology demonstrated that commuting matrices could be precisely crafted using call detail records, offering a cost-efficient alternative to conventional approaches [30].

Building on this, Iqbal et al. (2014) suggested formulating OD matrices by leveraging mobile phone call detail records alongside sparse traffic data. By analyzing CDRs from nearly 2.87 million users in Dhaka, Bangladesh, they juxtaposed time-stamped tower locations with caller identities. Their optimization-centric method, complemented by a microscopic traffic simulation platform, sought to fine-tune scaling factors aligning with real-world traffic data, particularly shining when high-quality travel surveys and traffic details are scant [31].

Subsequently, Alexander et al. (2015) introduced a method to infer daily travel patterns, drawing from a colossal dataset of over 8 billion anonymized mobile phone records. Collated over two spring months in 2010 from the Boston metropolitan area, these records transformed into clusters representing areas of user activity. The assumed nature of these locations, whether residential, occupational, or otherwise, was derived from observational patterns. Relying on city visit survey data, the team employed probabilistic inferences to pinpoint travel initiation times. Their methodology's robustness was validated against conventional survey data, attesting to its practicality [32].

Bhandari et al. (2018) unveiled an innovative pathway in [33], recognizing the capability of mobile call data to determine zonal population sizes and to calculate origin-destination journeys, challenging traditional data acquisition techniques. The team deployed dual strategies: analyzing regular commutes and interpreting sequential journeys, the latter being a mix of professional and personal travels. These data-driven methodologies, when benchmarked against conventional household

survey metrics, underscored the mobile-based approach's enhanced reliability and cost-effectiveness [33].

Venturing into the digital realm, Wang and Taylor (2016) harnessed Twitter's location data to infer origin-destination flows. They delineated both weekday and weekend travel patterns for an expansive metropolitan region, employing a specific algorithm for mobile location data. By implementing preliminary analysis on GPS data to circumvent locational inaccuracies, they emphasized the strategic potential of OD matrices. Such matrices could empower policymakers to optimize transportation systems, foresee future infrastructural needs, and enhance daily commutes [34].

Lastly, Pourmoradnasseri et al. (2019) illuminated a technique for retracing individual pathways using Call Details Records (CDR) and the Visitor Location Registry (VLR). Their dataset, encompassing over 600 million anonymized events from Estonian users, was vast. With the aid of a second-order Markov model, they revived the obscured narratives of mobile user routes. After meticulously mapping cell-to-cell user pathways, they crafted and subsequently evaluated the OD matrix for Estonia's primary cities against the country's train passenger data, affirming the method's specificity and precision [35].

User Profiling

Mobile data analysis has opened doors to a fascinating application: the classification of user profiles. While at a glance, a mobile phone user might seem to fall into broad categories such as a worker, student, or retiree, the richness of Call Detail Record (CDR) data offers a more nuanced characterization. By examining users' daily routines, frequented locations, time allocations, and connections (be they coworkers, friends, or family), the mobile network provides a canvas for detailed

user profiling. This review identifies two pivotal studies that shed light on user behaviors and patterns depending on the day of the week.

In a 2012 study, Jiang et al. delved into an activity-based travel survey from Chicago. The survey encompassed feedback from nearly 30,000 participants (across 10,552 families) collected over one or two days between January and February 2008. The vast dataset enabled the authors to scrutinize three critical behavioral dimensions. They deduced that based on activities, the populace could be segmented into eight groups for weekdays and seven for weekends. When juxtaposed with socio-demographic information, the ensuing clusters yield invaluable insights, particularly for urban planning, transportation strategies, and even considerations like emergency responses and contagion dynamics [36].

A later study in 2018 by Thuillier et al. unveiled human mobility patterns by examining an expansive dataset of over 800 million CDRs. The authors delineated users into six distinct profiles, such as "Resident Working in Zone" and "Weekend" traveler, emphasizing the temporal and spatial attributes of each category. Adopting an event-driven algorithm, they clustered individuals into 12 overarching weekly patterns. This method was particularly adept at addressing the inherent sparsity challenges in CDR data. Thuillier and the team posited that their innovative approach furnished significant depth and reliability to mobility analysis [37].

Group Movement Patterns

In urban settings, while individual mobility patterns can be observed, there is a distinct need to examine the collective movements and behaviors of groups. Such collective patterns are critical for detecting and understanding events such as large gatherings, escalating social protests, traffic disruptions, fires, and other incidents. Furthermore, some of these events have the potential to instigate significant

migration shifts, whether as internal displacements or as cross-border refugee movements. This section provides a literature review on methodologies for event detection, patterns of migration, and strategies for disaster preparedness.

Events Detection

In recent years, the analysis of mobile data has emerged as a potent tool for understanding crowd dynamics, especially during significant events in urban areas. Calabrese et al. (2010) delved into this realm by studying crowd movements during notable events in the Boston metropolitan area from July 30th to September 12th, 2009. Their dataset, comprising almost a million cell phone traces, associated specific user destinations with social events. Their research illustrated that the demographics of event attendees are strongly correlated with the event type. Such insights are invaluable for city planners, offering data-driven decisions on event management and strategies to mitigate congestion. The study underscored that proximity plays a role in event attraction and that similar events exhibit comparable attendee origins, enabling predictions on attendee demographics for future events [38].

Traag et al. (2011) ventured into the development of a methodology to pinpoint social events within vast mobile phone data pools. Their approach, rooted in a Bayesian location inference framework, offers a probabilistic perspective on user attendance at events. Harnessing 14 months of call data from approximately 5.75 million users in a European country, the authors revealed that deviations from routine movements signify event attendance. Such methodologies, they argue, are crucial for effective crowd management and timely detection of emergencies [39].

Focusing on Buenos Aires, Ponieman et al. (2013) employed mobile phone data to quantify the impact of significant social phenomena, like urban commutes or major

sporting events. Their model, predicated on analyzing most frequent past user locations, utilized five months of call data from a local operator. The study demonstrated the efficacy of mobile data in quantifying events, exemplified by their ability to estimate the attendance at soccer matches [40].

Hajdú-Szücs et al. (2018) provided insights into the potential of anomaly detection in mobile phone data analysis. Using a month's worth of anonymized data from a Hungarian service provider, they applied data mining and anomaly detection techniques. Their research highlighted the ability to infer the underlying reasons for observed human behaviors, marking a significant step towards potential anomaly and crowd detection [41].

In a similar timeframe, Pinter et al. (2018) devised a methodology to discern urban population patterns during major social events. Using CDR data, they contrasted average population densities with recent data for specific times and locations. Their approach, tested on data surrounding a 2017 public demonstration in Budapest, elucidated the intricacies of urban mobility during large-scale events [42].

Migration Patterns

Migration patterns have been a subject of increasing interest, particularly as researchers have harnessed mobile phone data to gain deeper insights. In a study by Williams et al. (2015), the authors critically evaluated existing mobile phone-based measures of mobility. They highlighted potential issues and introduced novel methods tailored to address them. Using anonymized Call Detail Records (CDRs) sourced from Rwanda's premier cellular phone service provider, they innovated new measures of mobility that cater to the spatial essence of human movement. These measures, by design, are detached from socio-contextual characteristics and are standardized for cross-regional and temporal comparability. Williams and his team

elucidated the potential applications of these metrics, suggesting they can enhance the understanding of human behavior at the micro-level, societal structures at the macro-level, and the consequent shifts over time [43].

Taking a closer look at urban mobility, Shi et al. (2017) embarked on a comprehensive analysis of human movement in a densely populated district of Beijing. Their dataset comprised an impressive 2.38 million anonymized CDRs representing 10,000 users sampled in June 2013. Preliminary findings painted a picture of Beijing's residents leading exceptionally busy lives, with their mobility patterns more frenetic than their counterparts in Portugal and the Ivory Coast. The study utilized a convex hull analysis to reveal that an individual's daily trajectory area adheres to a power-law distribution. Interestingly, the study found only a slight correlation between call frequency and parameters like travel distance, movement radius, and the aforementioned convex hull area. Furthermore, their research unearthed a cyclic pattern in daily migrations, identifying a consistent return point between 13:00 and 15:00 [44].

In a timely study, Jia et al. (2020) turned to mobile data to understand the population outflows from Wuhan, China, in relation to the early stages of the COVID-19 pandemic. Harnessing data from over 11 million mobile counts between January 1 and January 24, 2020, the team constructed a 'risk source' model that operated on spatio-temporal parameters. This model projected the likely distribution of COVID-19 cases, pinpointing high-risk areas of transmission even in the pandemic's nascent stages. Such an approach provides policymakers globally with a tool to perform swift risk assessments, enabling them to judiciously allocate resources during outbreaks [45].

These studies underscore the invaluable role of mobile data in advancing our understanding of migration patterns and their wider societal implications.

Disaster Planning

In recent years, the utilization of mobile data has become integral to understanding human movement during and after significant disaster events. Lu et al. (2012) embarked on an insightful analysis surrounding the catastrophic Haiti earthquake of January 12, 2010. They assessed the movements of 1.9 million mobile phone users over an extended period, spanning 42 days before to 341 days after the earthquake. The dataset, courtesy of the mobile operator Digicel, unveiled a notable increase in travel distances and the magnitude of individuals' movement trajectories post-disaster. This study shed light on the tangible alterations in mobility patterns experienced by residents in the aftermath of such a profound event [46].

Moving to the Far East, Song et al. (2014) assembled an extensive database documenting human mobility, with the GPS records of 1.6 million users over a year. Their objective was to delve into human emergency behavior following the dual tragedies of the Great East Japan Earthquake and the subsequent Fukushima nuclear incident. Through empirical analysis, the researchers discerned correlations between human mobility post-disasters and their regular movement patterns. Additionally, they identified that factors such as social ties, the severity of the disaster, infrastructural damage, governmental aid, media coverage, and the migration of large populations significantly influence behavior during crises. Building on these findings, they formulated a behavior model that captures these dimensions, asserting that human behavior in disaster contexts is more predictable than traditionally believed [47].

Lu and team returned in 2016 with another seminal work, this time spotlighting Bangladesh. They explored the potential of mobile network data in comprehending movement dynamics during and subsequent to severe weather events. Drawing from mobility trajectories of six million users, they scrutinized short-term mobility aspects during and post the Cyclone Mahasen of May 2013. Such granular features of human movement, usually elusive in conventional survey methodologies, were distinctly captured. Their analysis underscored the capability of mobile data in elucidating intricate connections between key migratory facets on a grand scale. This contribution has since been acknowledged as a cornerstone in the realms of human migration research, especially in the context of climate change [48].

In summary, these studies embody the transformative role of mobile data in enhancing the comprehension and predictability of human behavior during disastrous events.

Epidemic Spread

The global outbreak of COVID-19 and the subsequent worldwide quarantines to curb its spread highlighted the significance of studying human mobility. While the initial implications of a pandemic were not immediately understood by many, real-time virus tracking illuminated its impact on public health. Data gathered in each country became instrumental for devising strategies to curb the disease's progression, given that public health remains a paramount concern for governments worldwide. This section spotlights four research endeavors that delve into disease spread, ranging from the 2011 H1N1 outbreak to the more recent COVID-19 pandemic.

In 2011, Frias-Martinez et al. proposed an agent-based system to model virus dissemination. Drawing from call detail records, they investigated the spread of the H1N1 virus in Mexico during its 2009 outbreak and evaluated the influence of

government interventions on the epidemic. The system is anchored in two main components: (1) agents structured around call detail records and (2) a discrete event simulator (DES) that employs these agent models to project viral spread over time. Notably, the research illustrated that mobility restrictions implemented by the government curtailed the peak number of infections by 10% and delayed the peak of the pandemic by two days [49].

Later, in 2015, Bengtsson et al. examined if mobile operator data could forecast the early spatial evolution of the 2010 cholera outbreak in Haiti. The study leveraged data from 2.9 million anonymized mobile phone SIM cards alongside highly detailed case data. They contrasted two gravity models of population mobility, both calibrated using comprehensive retrospective epidemic data available only post-epidemic. The ensuing analysis showed a robust correlation between the risk of an area witnessing an outbreak within a week and the mobile phone-derived infectious pressure metrics. Such insights underscore the promise of mobile data in bolstering preparedness and responses during cholera outbreaks, with implications for other infectious disease containment efforts [50].

Fast forward to 2020, Fang et al. assessed the ramifications of the Wuhan lockdown on January 23, 2020, especially regarding the containment and delay of the 2019-nCoV spread. To discern the lockdown's influence on human mobility from other factors such as panic and the Spring Festival, they employed difference-in-differences (DID) estimations using inter-city migration data sourced from Baidu Migration. Their research affirmed that stringent social distancing measures across 98 Chinese cities outside the Hubei province effectively minimized the effects of incoming populations from Hubei's epicenter on the destination cities' 2019-nCoV spread. These findings hold significant implications for global pandemic mitigation efforts [51].

A year later, in 2021, Ayan et al. harnessed cellular network usage data from Rio de Janeiro to fathom and model human mobility patterns during the COVID-19 pandemic. Their dataset spanned user aggregate and individual connection data, encompassing approximately 1400 cellular antennas in Rio de Janeiro from March 1 to July 1, 2020. The analysis highlighted that, intriguingly, human mobility surged even before official lockdown relaxations, a trend persisting post-lockdown. Particularly, Friday witnessed enhanced mobility, suggesting people possibly loosening personal restrictions and partaking in social activities. The researchers also introduced an interactive tool, enabling individuals and officials to gain insights into mobility patterns and local COVID-19 case counts [52]. In essence, understanding human mobility can be a potent tool in anticipating and managing disease outbreaks.

The literature review shed light on the role of digital technologies in exploring human mobility patterns within urban contexts. Through cellular data, researchers have endeavored to create models offering deeper insights into these mobility dynamics.

It's crucial to note that the bulk of the literature primarily examined call detail records for specific applications within confined geographical regions. Consequently, such findings may not be representative of broader city dynamics or the routines of its inhabitants.

To truly grasp city dynamics, data spanning a wide geographical scope over an extended duration is imperative. Such data enables capturing diverse user profiles and understanding their daily patterns. For the upcoming study, there is access to a comprehensive data set detailing exact locations of individuals across various U.S. cities. This rich data source offers a unique opportunity to extract numerous parameters and discern patterns. The study provides a platform to compare findings with those in existing literature, particularly where synthetic or limited CDR data

sets were employed. Furthermore, it facilitates comparisons of human mobility dynamics across cities within the same or different countries. These analyses will highlight both similarities and disparities among the cities in focus. Additionally, the models devised can prove instrumental in informing decisions related to intelligent transport system deployment, urban planning, anomaly detection in cities, and tracking epidemics or pandemics with potential public health implications.

Chapter 3
Dataset Description

This research utilizes a dataset that comprises commercially available location information, predominantly GPS coordinates, from mobile devices. This data originates from X-Mode Social. According to X-Mode Social's website, the X-Mode SDK integrates into over 300 applications across diverse categories, including gaming, travel, and dating. This integration facilitates the collection of data from over 50 million active users each month, with locations being shared every 5 to 7 minutes [53]. The gathered data can be attributed to a combination of GPS, Bluetooth signals detected by beacons, and Wi-Fi router signals. Notably, users retain the discretion to disable this data-sharing via their smartphone settings [54].

The dataset, spanning from October 1st to October 30th, 2020, provides daily location data from mobile devices throughout the United States. Three significant subsets of this dataset have been extracted and analyzed in this research representing different geographical domains:

1. The first subset focuses on the Houston MSA in Texas, which stretches across about 9,444 square miles. Home to roughly 7.2 million residents, the dataset contains records from 142,132 distinct phones, which translates to around 1.97% of the region's population [55]. The area is depicted in Figure 3-1.
2. The second subset encompasses location data from the NY-NJ-PA MSA, spreading over 6,684 square miles and supporting a population of approximately 19,768,458. From this region, records of 278,742 unique devices were analyzed, approximating 1.4% of the region's total inhabitants [56]. A visual representation is displayed in Figure 3-2.

3. The third subset compiles location data from the fifteen most populous MSAs in the U.S., inclusive of the aforementioned NY-NJ-PA and Houston MSAs. Relevant details, such as population figures, individual mobile device counts, and geographically binned data, are itemized in Table 3-1.

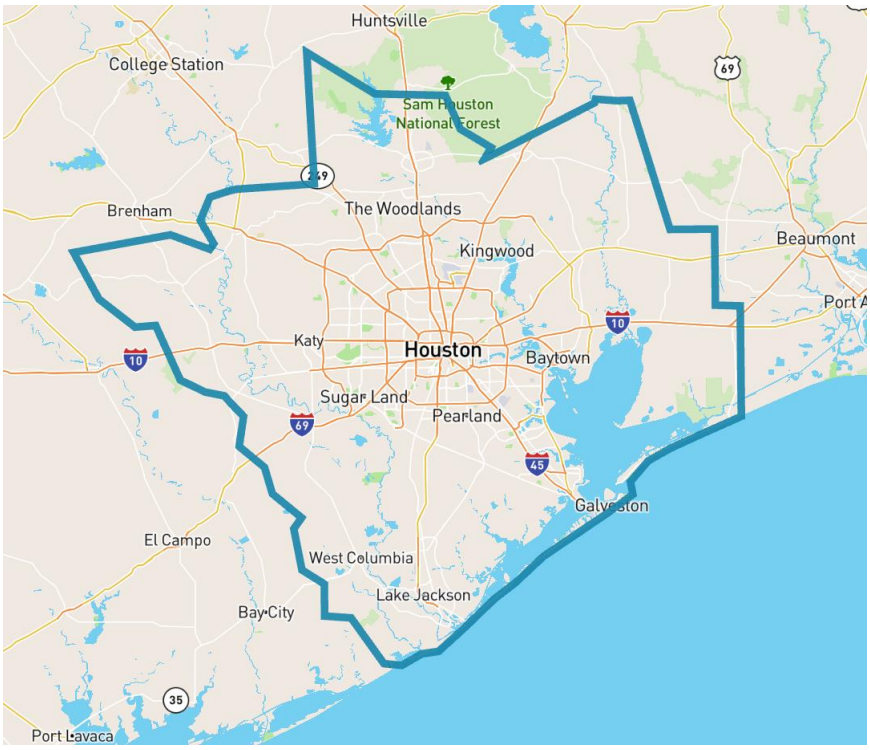

Figure 3-1: Houston Metropolitan Statistical Area.

Figure 3-2: NY-NJ-PA Metropolitan Statistical Area.

All datasets used in this research have been geographically binned employing Uber's H3 Hexagonal Hierarchical Spatial Index [57]. This system offers 15 predefined resolution levels that define the radius of each hexagonal bin. For this research, three specific H3 resolution levels have been chosen and are consistently used throughout the dissertation, as illustrated in Table 3-1. The number of locations for the 15 studied MSAs, listed in Table 3-2, is determined based on the h8 resolution level, corresponding to an approximate hexagonal bin radius of 531 meters. The population percentages cited for each MSA stem from the U.S. Census Bureau's 2020 population estimates [58].

The raw dataset schema is comprehensive, enumerating columns such as

"**advertiser_id**", "**platform**", "**location_at**", "**latitude**", "**longitude**", "**altitude**", "**horizontal_accuracy**", "**vertical_accuracy**", "**heading**", "**speed**", "**ipv_4**", "**ipv_6**", "**final_country**", "**user_agent**", "**background**", "**publisher_id**", "**wifi_ssid**", "**wifi_bssid**", "**tech_signals**", "**carrier**", "**model**", "**venue_name**", "**venue_category**", and "**dwell_time**".

It's worth noting that "**location_at**" may also manifest as "**timestamp**", both represented in Coordinated Universal Time (UTC) format.

For the scope of this research, the primary columns of interest are:

- *Advertiser_id*: A distinct identifier for each mobile unit, allocated by the operating system of the device. This is not tied to a phone number, safeguarding user anonymity. Nevertheless, this number facilitates the tracking of a device's trajectory over time.
- *Location_at (timestamp)*: Corresponding to the UNIX time of the data point. UNIX time denotes elapsed seconds from the fixed reference of January 1, 1970, 00:00:00, with this dataset detailing it in seconds.
- *Latitude*: The north-south positional metric of the device, expressed in decimal degrees, with positive figures denoting the northern hemisphere and negative ones the southern hemisphere.
- *Longitude*: The east-west coordinate of the device, illustrated in decimal degrees. Positive coordinates are indicative of locations in the eastern hemisphere, while negative ones denote the western hemisphere.

This research emphasizes the importance of data privacy by concentrating solely on anonymized datasets, extracting profound insights from this granular, expansive data.

Table 3-1: Uber H3 Resolution Levels

Notation	Resolution Level	Radius (m)
h7	7	1406
h8	8	531
h9	9	200

Table 3-2: Population, Unique Phones, And Locations

Metropolitan Statistical Area (MSA)	**Population (2020)**	**Unique Phones**	**Locations**
New York-Newark-Jersey City, NY-NJ	20,140,470	278,742	31,511
Los Angeles-Long Beach-Anaheim, CA	13,200,998	74,074	10,964
Chicago-Naperville-Elgin, IL-IN	9,618,502	147,635	25,113
Dallas-Fort Worth-Arlington, TX	7,637,387	173,883	29,723
Houston-Pasadena-The Woodlands, TX	7,122,240	142,132	26,276
Washington-Arlington-Alexandria, DC-VA-MD-WV	6,385,162	71,652	22,176
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	6,245,051	87,404	16,767
Atlanta-Sandy Springs-Roswell, GA	6,089,815	115,796	32,029
Miami-Fort Lauderdale-West Palm Beach, FL	6,138,333	86,494	12,341
Phoenix-Mesa-Chandler, AZ	4,845,832	70,848	20,294
Boston-Cambridge-Newton, MA-NH	4,941,632	51,309	12,780
Riverside-San Bernardino-Ontario, CA	4,599,839	31,313	20,176
San Francisco-Oakland-Fremont, CA	4,749,008	23,100	6,285

Detroit–Warren–Dearborn, MI	4,392,041	71,119	13,988
Seattle-Tacoma-Bellevue, WA	4,018,762	50,544	13,432

Data Preprocessing

Due to the vast size of the raw dataset, this research employs the Hadoop Ecosystem, an integrated suite of services optimized for storing and distributing large datasets across multiple nodes within a cluster. Data processing is executed using Apache Spark [59], a framework that is commonly paired with the Hadoop Ecosystem to perform large-scale data ETL (Extract, Transform, and Load) tasks. The processed data is stored in the Hadoop Distributed File System (HDFS) in Optimized Row Columnar (ORC) format.

The raw sample data underwent essential preprocessing steps to enhance data quality and refine the accuracy of the results. These steps involved cleaning and transforming the data to make it suitable for the planned analyses. The steps can be summarized as follows:

1. *Timestamp Adjustment*: The raw data timestamps, originally based on UTC, are adjusted to match the local time of the studied MSA.
2. *Geographical Binning*: Data is grouped geographically into bins according to the three H3 resolution levels provided in Table 3-1. Every H3 hexagonal bin is assigned a unique location ID and center coordinates. These values replace the actual GPS coordinates for each data record.
3. *Time Annotations*: The dataset is augmented to include "**start-time**," "**end-time**," and "**stay-duration**" for every individual on a daily basis throughout the study duration. Specifically:
    - A new location for an individual is noted with a timestamp marked as the "**start-time**."

- If the record isn't the first location entry for the day for the individual, that timestamp also serves as the "**end-time**" for the previous location.
- If the individual doesn't record a new location by the day's end, the day's final record for the current location is designated the "**end-time**."
- The "**stay-duration**" for each record is computed as the time difference between the "**start-time**" and "**end-time**."

4. *Data Filtering*: The dataset is filtered to only include those individuals who are present every day of the month and have at least 24 records daily.

Subsequent analyses in later chapters are conducted on daily, weekly, and monthly scales, incorporating various geographical bin resolutions. The investigation covers a period of four weeks, with distinctions made between weekdays and weekends. From the dataset, the Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) for mobility parameters are derived for each time scale.

Chapter 4
Mobility Parameters: Definition and Analyses Results

Definition

In this study, we consider a set of key parameters related to human mobility, emphasizing the spatial aspects of movement. Subsequently, definitions for each parameter are Provided below. For a comprehensive analysis, each parameter is evaluated across three distinct observation periods or time scales: daily, weekly, and monthly.

A) Number of Visited Locations (N_LOC)

A location refers to an area where an individual lives, works, or travels. In this paper, a hexagonal bin is used to represent a distinct location. The total number of bins visited within a given time frame is indicated as N_LOC [3], [19], [20]. A location is considered visited when an individual spends a specified duration of time within that location. Multiple visits to the same location within the observation period are each included in the overall count of N_LOC.

B) Number of Unique Visited Locations (N_ULOC)

A person's daily trajectory usually includes visits to different locations. The number of visited locations may include the same locations visited at various times during the observation period. The set of unique visited locations is a subset of the visited locations. A unique location counts only once, regardless of the number of visits. The parameter N_ULOC is the number of unique visited locations [3], [19], [20].

C) Number of Significant Locations (N_SIG)

Within a predetermined observation period, individuals are likely to visit a range of locations, among which some are of greater importance than others. To assess the importance of a particular location, a minimum duration of stay is defined. A location achieves significance if an individual remains at that location for at least four hours within the specified observational timeframe. The term N_SIG is used to represent the count of the number of significant locations visited by individuals based on this time requirement.

D) Radius of Gyration (R_GYR)

The Radius of Gyration (R_GYR) is a measure of an individual's mobility area size. It represents the greatest distance traveled by an individual from its mobility center of mass [19], [20], [60].

Mathematical definition: Given a list of recorded locations ID's loc_i and stay duration t_i for individual k, $(loc_i, t_i)_{i=1,\dots,N_k}$, denote with $c_i = (x_i, y_i)$ the coordinates of location i (longitude and latitude), the "Mobility Center of Mass" (MCM) is defined as:

$$c_m = \begin{pmatrix} x_{cm} \\ y_{cm} \end{pmatrix} \quad (1)$$

Where:

$$x_{cm} = \frac{1}{\hat{t}_k} \sum_{i=1}^{N_k} x_i t_i \quad (2)$$

$$y_{cm} = \frac{1}{\hat{t}_k} \sum_{i=1}^{N_k} y_i t_i \quad (3)$$

And,

$$\hat{t}_k = \sum_{i=1}^{N_k} t_i \quad (4)$$

Finally, R_GYR is taken to be the maximum of the distances between any location for individual k and the MCM.

$$R\_GYR = max_{i=1}^{N_k}\big(D(c_i, c_m)\big) \quad (5)$$

It is important to note that this calculation considers all locations, irrespective of whether they are classified as visited or not. $D(\cdot)$ calculates the geographical straight-line distance between the i-th location and the MCM.

E) Distance Traveled (D_TRAV)

D_TRAV is defined as the total linear distance traveled by an individual at a given time scale [19], [20], [61].

Mathematical definition: Similar to R_GYR, for each individual k, we denote with $c_i = (x_i, y_i)$ the coordinates of location i. D_TRAV is the defined as:

$$D = \sum_{i=1}^{N_k - 1} D(c_i, c_{i+1}) \quad (6)$$

Where $D(\cdot)$ calculates the distance between two locations.

F) Travel Time Percentage (T_TP)

In this study, T_TP is used to describe the portion of time individuals spend in transit relative to the total observation period. While it's previously noted that individuals spend a significant amount of time at particular locations, the time spent

traveling between these key locations should not be overlooked. Consequently, understanding the ratio of travel time to the time spent at visited locations is critical.

The travel time percentage T_TP is calculated as:

$$T\_TP = \frac{T_{total} - \sum_{i=1}^{N_k} t_i}{T_{total}} \times 100\% \quad (7)$$

Where T_{total} denotes the total observed time for an individual k within a given time frame, and t_i represents the duration, in seconds, spent at location i.

In summary, the parameters under consideration have been developed to capture the mobility patterns exhibited by a group of individuals across various time scales. The initial three parameters assess visit behavior (N_LOC, N_ULOC, and N_SIG), while the remaining parameters consist of two parameters that evaluate displacement behavior (R_GYR and D_TRAV), and finally, T_TP assesses spatio-temporal behavior.

Sampling Thresholds

Sampling thresholds in this study are applied to mobility parameters in two distinct Metropolitan Statistical Areas (MSAs): Houston and NY-NJ-PA. Two primary sampling thresholds are examined. First, the research uses the Uber H3 indexing system for geographic binning, analyzing data across different hexagonal geospatial resolutions. This approach aims to discern how variations in spatial granularity can influence the understanding of mobility patterns. The second sampling strategy involves determining the minimum time that qualifies a location as a 'visit', examining the effects of different stay durations on movement patterns.

For the Houston MSA, the analysis focuses on four mobility parameters: N_LOC, N_ULOC, R_GYR, and D_TRAV. The data is categorized using three specific

geographic binning resolutions: h7, h8, and h9, as referenced in Table 3-1. Locations with stay durations exceeding 10, 15, or 20 minutes are classified as "visited" for parameters N_LOC and N_ULOC.

In the New York-Newark-Jersey City (NY-NJ-PA) MSA, the scope is expanded to six mobility parameters, introducing N_SIG and T_TP alongside the initial four. The same geographic binning resolutions are employed. Besides the visitation criteria used in the Houston study, a location is considered "significant" if an individual's stay lasts at least four hours within the observation timeframe.

For both MSAs, it is important to note that displacement parameters, R_GYR and D_TRAV, do not use stay duration thresholds. The reason is straightforward: these parameters specifically measure displacement and are not associated with visit behavior, making the consideration of stay durations irrelevant for these parameters. Both studies encompass daily, weekly, and monthly observation intervals.

The subsequent sections will present the results for each mobility parameter within the context of the respective MSAs.

Houston MSA

In Chapter 3 – Data Preprocessing, the detailed preprocessing steps for handling the extensive raw dataset are presented. Upon preprocessing, the analysis is conducted for each geographical binning sampling threshold (h7, h8, and h9) across daily, weekly, and monthly scales. For daily examinations, weekdays and weekends receive separate attention. For the parameters N_LOC and N_ULOC, stay duration thresholds of 10, 15, and 20 minutes are implemented. Any records with "**stay-durations**" below these thresholds are excluded, effectively omitting all transitional locations. Estimates of the Probability Mass Function (PMF) and

Cumulative Distribution Function (CDF) for the four mobility parameters are derived from the dataset for each time scale. The total number of unique IDs and Locations for Houston MSA, following these processes, is presented in Table 4-1. In this research, each distinct bin represents a unique location.

Table 4-1: Number of IDs and Locations Per Res. Level. (Houston MSA)

	h7	**h8**	**h9**
IDs	142132	142132	142132
Locations	4185	26276	145397

A) Number of Visited Locations (N_LOC)

Daily Analysis

Table 4-2 shows the daily average mean and standard deviation for different sampling thresholds. It may be observed that, for a hexagonal bin resolution of h7 and a stay duration threshold of 10 minutes, an individual in the Houston MSA area visits an average of 5.8 locations during weekdays and 5.4 locations during weekends. The corresponding standard deviations are approximately 4.5 and 4.3, respectively. When the duration threshold is increased while maintaining the same bin size, the average number of visited locations is reduced. This behavior is expected since the stay duration threshold determines whether a location is considered visited or not. By increasing the stay duration threshold, fewer locations will be considered visited.

When the hexagonal bin resolution is increased to h8 and a stay duration threshold of 10 minutes is applied, the average number of visited locations is higher compared to resolution h7. The average number of visited locations is around 6.7 during weekdays and 6.3 locations during weekends. The average standard deviation is also

higher, at around 5.3 and 5.2, respectively. This is because higher resolutions are associated with smaller bin sizes. Table 4-1 shows that when the Houston MSA area is binned using resolution h7, there are 4185 unique locations (bins) compared to 26276 unique locations for resolution h8. Thus, observing a higher average number of visited locations is easily explainable. The same behavior is observed when increasing the resolution level to h9. Additionally, for both resolutions h8 and h9, increasing the stay duration threshold still results in a lower average number of visited locations and a lower standard deviation.

Figure 4-1 shows the daily average for the N_LOC over the month for the proposed sampling thresholds. The results are consistent across the various sampling thresholds. Additionally, Fig. 4-1 indicates that regardless of the hexagonal bin resolution and the stay duration threshold, the average number of places visited during the week is slightly higher on weekdays, with Friday having the highest average. In contrast, the lowest mobility during the week appears to be on Sundays.

An illustrative set of Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) for the number of visited locations on a weekday (10/01/2020) is shown in Fig. 4-2. PMFs and CDFs of similar shape are obtained for all other days of the month. The curves shift slightly as a result of sampling thresholds, but the shape of the curves remains the same. The N_LOC PDF seems to peak at three (3) visited locations and is always less than 30 on a daily time scale. Moreover, a dip in N_LOC can also be noticed at two (2) visited locations, indicating that the number of individuals observed in only two locations during the day is low. This result is rather expected. Let us assume that the first location is the home location, and the second location is a grocery store. The common scenario is that the individual would return home before the end of the day. Therefore, it is expected that the number of individuals visiting only two locations is small.

Table 4-2: N_LOC & N_ULOC: Daily Average Mean and Standard Deviation for Different Sampling Thresholds

	Bin Res.	**h7**			**h8**			**h9**		
	Time Scales	**10 Min**	**15 Min**	**20 Min**	**10 Min**	**15 Min**	**20 Min**	**10 Min**	**15 Min**	**20 Min**
N_LOC Avg Mean	**Weekday**	5.8	5.0	4.6	6.7	5.8	5.2	8.3	7.0	6.2
	Weekend	5.4	4.7	4.3	6.3	5.4	4.8	7.7	6.5	5.7
	Week	33.8	28.9	25.9	40.3	34.0	30.0	51.6	42.7	37.0
	Month	142.5	121.7	108.7	170.5	143.4	126.3	219.2	181.1	156.5
N_LOC, Avg StdDev	**Weekday**	4.5	3.6	3.0	5.3	4.2	3.5	6.6	5.2	4.3
	Weekend	4.3	3.4	2.9	5.2	4.1	3.5	6.5	5.2	4.3
	Week	24.5	19.5	16.5	30.3	24.0	20.1	39.2	31.1	25.8
	Month	98.7	78.6	66.5	122.6	97.3	81.4	159.3	126.8	105.1
N_ULOC Avg Mean	**Weekday**	3.4	2.9	2.7	3.7	3.2	2.9	4.2	3.5	3.1
	Weekend	3.3	2.9	2.6	3.6	3.1	2.8	4.1	3.4	3.0
	Week	10.8	9.1	8.1	13.5	10.9	9.4	16.2	12.6	10.5

	Bin Res.	h7			h8			h9		
	Time Scales	**10 Min**	**15 Min**	**20 Min**	**10 Min**	**15 Min**	**20 Min**	**10 Min**	**15 Min**	**20 Min**
	Month	25.4	21.5	19.0	35.4	28.5	24.3	46.1	35.4	29.0
N_ULOC, Avg StdDev	**Week-day**	2.2	1.7	1.5	2.5	1.9	1.6	2.8	2.2	1.8
	Week-end	2.1	1.7	1.5	2.4	1.9	1.6	2.7	2.1	1.8
	Week	6.9	5.5	4.7	8.6	6.6	5.5	10.3	7.7	6.2
	Month	16.8	13.4	11.5	23.2	17.5	14.6	29.9	21.7	17.4

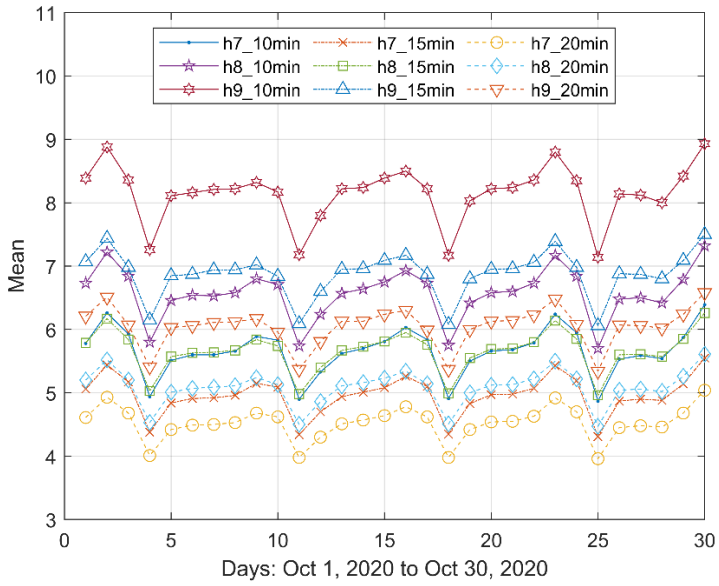

Figure 4-1: Mean of N_LOC between 10/01/2020 and 10/30/2020.

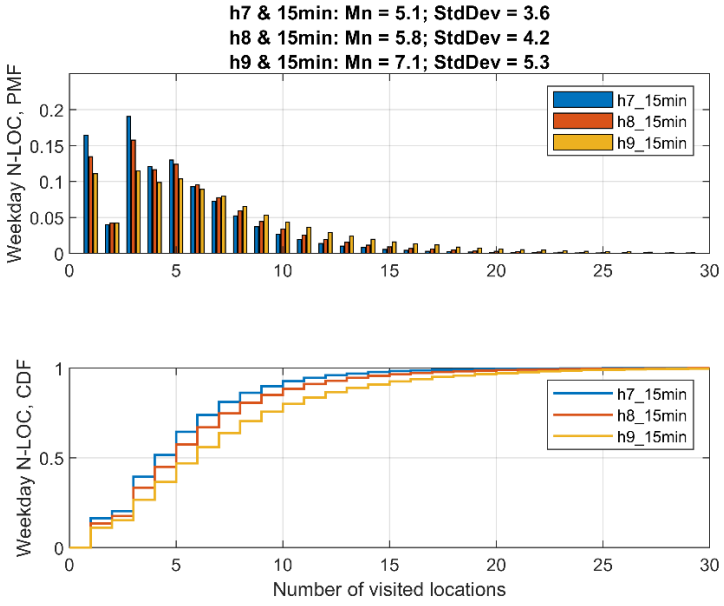

Figure 4-2: Estimated PMF and CDF for N_LOC weekday: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/01/2020).

Weekly Analysis

Table 4-2 presents the weekly average mean and standard deviation for the different sampling thresholds. It can be observed that for hexagonal bin resolution h7 and a stay duration threshold of 10 minutes, an individual in the Houston MSA area visits an average of 33.8 locations per week with an average standard deviation of around 24.5. Similar to the daily time scale, the average number of visited locations is lower when the duration threshold is increased while keeping the same bin size. On the other hand, increasing the hexagonal bin resolution to h8 and applying a stay duration threshold of 10 minutes results in a higher average number of visited locations per week than resolution h7. The average number of visited locations per week is around 40.3, and the average standard deviation is also higher at around 30.3.

The first week of the month (October 1 to October 7) is selected in this analysis to show the estimated PMF and CDF for N_LOC. The estimated PMF and CDF are shown in Figure. 4-3 and Figure 4-4. Figure 4-3 presents the PMF and CDF for resolution levels h7, h8, and h9 with a stay duration threshold of 15 minutes. Figure 4-4 shows the PMF and CDF for resolution level h8 with a stay duration threshold of 10, 15, and 20 minutes. Similar to the daily time scale, Figure 4-3 and Figure 4-4 show that the estimated PMF curves obtained from the different sampling thresholds are quite similar. However, there are slight differences in the peaks of N_LOC for each sampling threshold. Additionally, N_LOC is consistently smaller than 150 when viewed on a weekly time scale.

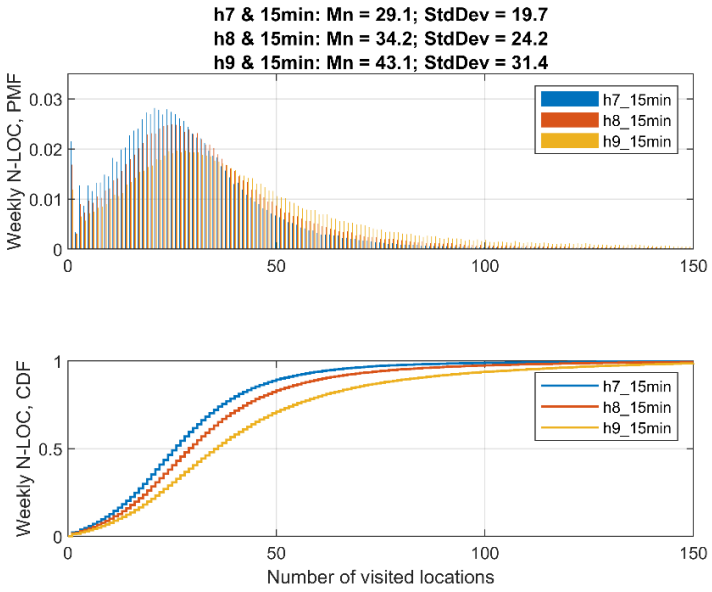

Figure 4-3: Estimated PMF and CDF for N_LOC weekly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 1 to October 7).

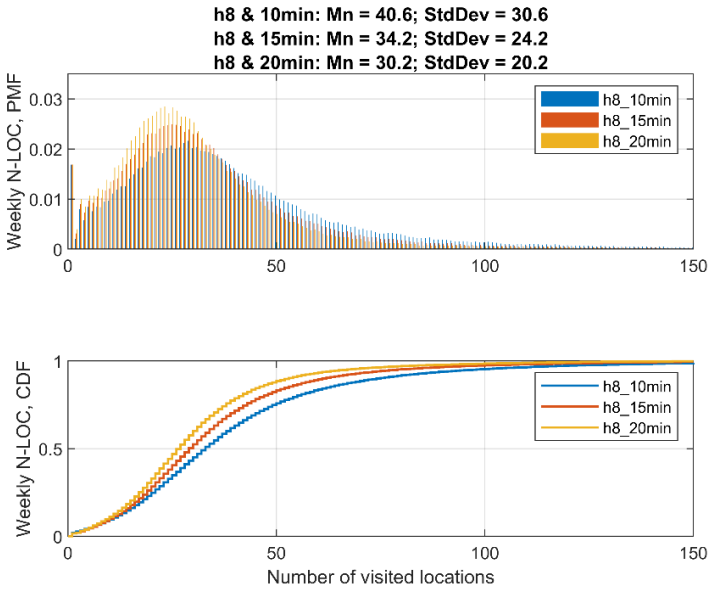

Figure 4-4: Estimated PMF & CDF for N_LOC weekly: Res. h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 1 to October 7).

Monthly Analysis

Table 4-2 shows the monthly average mean and standard deviation for various sampling thresholds. For hexagonal bin resolution h7 and a stay duration threshold of 10 minutes, individuals in the Houston MSA area visit an average of 142.5 locations per month, with a standard deviation of approximately 98.7. As with lower time scales, increasing the duration threshold while keeping the same bin size leads to a lower average number of visited locations. Similarly, a higher hexagonal bin resolution results in a higher average number of visited locations.

Figure 4-5 shows the PMF and CDF for resolution levels h7, h8, and h9 with a stay duration threshold of 15 minutes. Figure 4-5 also shows that the estimated PMF curves for the different sampling thresholds are similar for different spatial resolutions. However, there are slight differences in the N_LOC peak and upper limit for each of the thresholds.

B) Number of Unique Visited Locations (N_ULOC)

Daily Analysis

The daily means and standard deviations for the different sampling thresholds are shown in Table 4-2. It may be observed that for hexagonal bin resolution h7 and a stay duration threshold of 10 minutes, an individual in the Houston MSA area visits an average of 3.4 unique locations during weekdays and an average of 3.3 locations during weekends. The standard deviation is around 2.2 and 2.1 for weekdays and weekends, respectively. When the duration threshold is increased while keeping the same bin size, the average number of unique visited locations is reduced to an average of 2.9 locations for both weekdays and weekends.

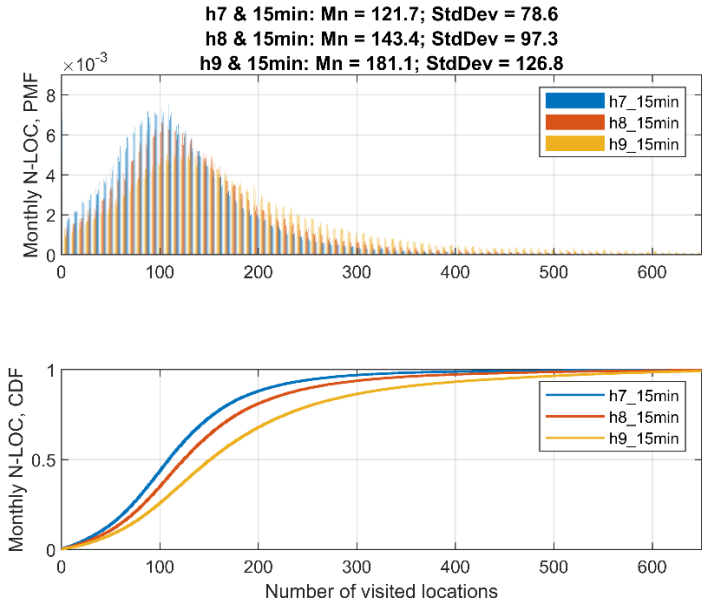

Figure 4-5: Estimated PMF and CDF for N_LOC monthly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 2020).

The average standard deviation is also lower, at around 1.7 for both weekdays and weekends. The same observation still applies when the stay duration is increased to 20 minutes. This behavior is expected. By increasing the hexagonal bin resolution to h8 and applying a stay duration threshold of 10 minutes, which is similar to N_LOC, the average number of unique visited locations is higher compared to resolution h7. The average number of unique visited locations is around 3.7 during weekdays and an average of around 3.6 locations during weekends. The average standard deviation is also higher, at around 2.5 and 2.4 for weekdays and weekends, respectively, which is consistent with the observations from N_LOC. The same behavior is also observed when increasing the resolution level to h9. Additionally, for both resolutions h8 and h9, increasing the stay duration threshold still results in a lower average number of unique visited locations and lower standard deviations.

Figure 4-6 shows the average number of unique visited locations (N_ULOC) at a daily scale for the proposed sampling thresholds. Similar to N_LOC, it is evident that the results are quite consistent across the different sampling thresholds. Additionally, Figure 4-6 reveals that regardless of the hexagonal bin resolution and the stay duration threshold, the average number of places visited during the week is slightly higher on weekdays, with Friday having the highest average. The lowest mobility during the week appears to be on Sundays. The estimated PMFs and CDFs are shown in Figure 4-7 and Figure 4-8. Figure 4-7 shows the PMF and CDF for resolution levels h7, h8, and h9 with a stay duration threshold of 15 minutes. Conversely, Figure 4-8 shows the PMF and the CDF for resolution level h8 with stay duration thresholds of 10, 15, and 20 minutes. It is evident from these figures that the estimated PMF curves obtained from the different sampling thresholds are quite similar. Moreover, N_ULOC remains smaller than 15 on the daily time scale.

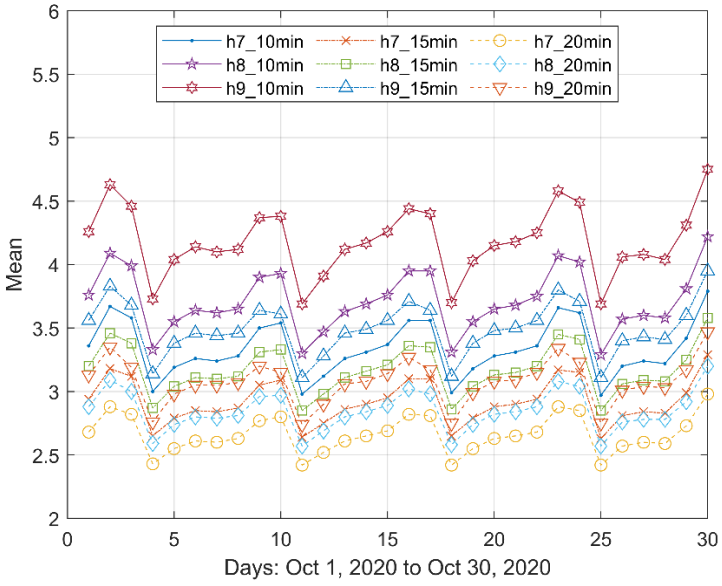

Figure 4-6: Mean of N_ULOC between 10/01/2020 and 10/30/2020.

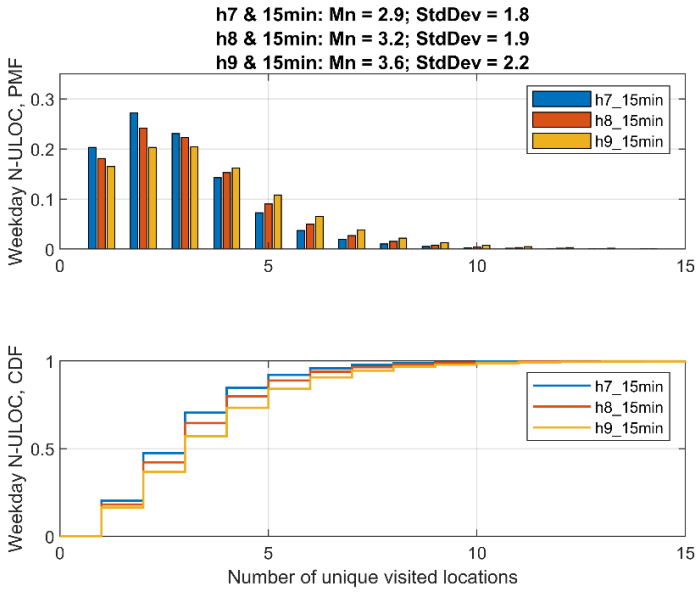

Figure 4-7: Estimated PMF and CDF for N_ULOC weekday: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/01/2020).

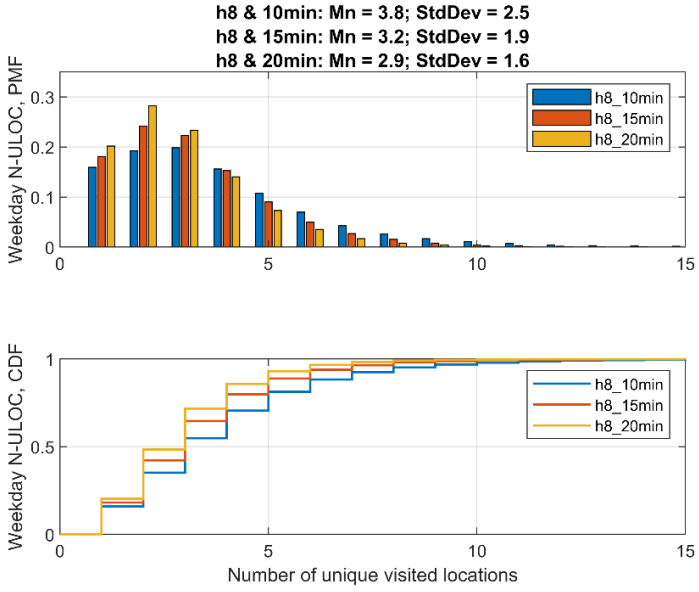

Figure 4-8: Estimated PMF and CDF for N_ULOC weekday: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/01/2020).

Weekly Analysis

Table 4-2 presents the weekly average mean and standard deviation for the proposed sampling thresholds. For hexagonal bin resolution h7 and a stay duration threshold of 10 minutes, an individual in the Houston MSA area visits an average of 10.8 unique locations per week, with an average standard deviation of around 6.9. Similar to the daily time scale, increasing the duration threshold while keeping the same bin size leads to a lower average number of visited locations. Furthermore, by increasing the hexagonal bin resolution to h8 and applying a stay duration threshold of 10 minutes, it is evident that the average number of unique visited locations per week is higher compared to resolution h7.

The average number of visited locations per week is around 13.5, and the average standard deviation is also higher, at around 8.6. The first week of the month (October 1 to October 7) is selected to illustrate estimated PMFs and CDFs for the number of unique visited locations. The estimated PMFs and CDFs are shown in Figure 4-9 and Figure 4-10. Figure 4-9 presents the PMFs and CDFs for resolution levels h7, h8, and h9 with a stay duration threshold of 15 minutes. Figure 4-10 shows the PMFs and CDFs for resolution level h8 with a stay duration threshold of 10, 15, and 20 minutes.

Similar to the daily time scale, it is seen that the estimated PMF curves obtained from the different sampling thresholds are very similar. However, each sampling threshold shows slightly different peaks for N_ULOC, and it can be observed that N_ULOC is consistently smaller than 50 on a weekly time scale.

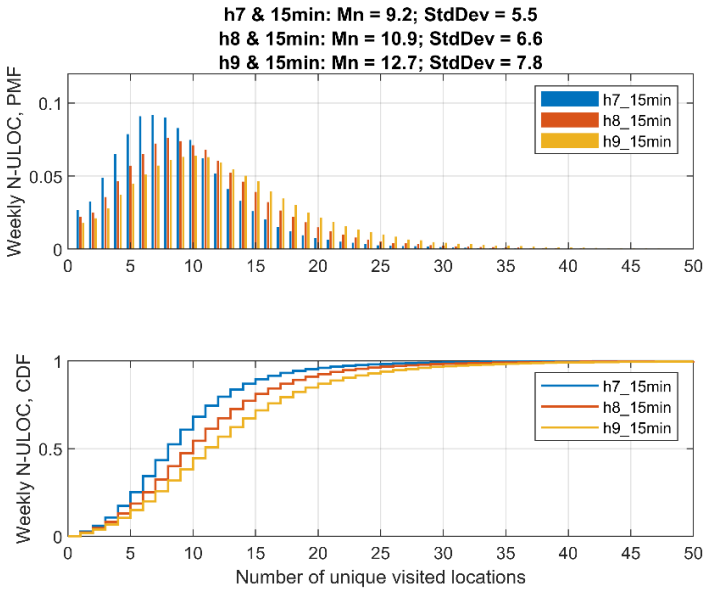

Figure 4-9:Estimated PMF and CDF for N_ULOC weekly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 1 to October 7).

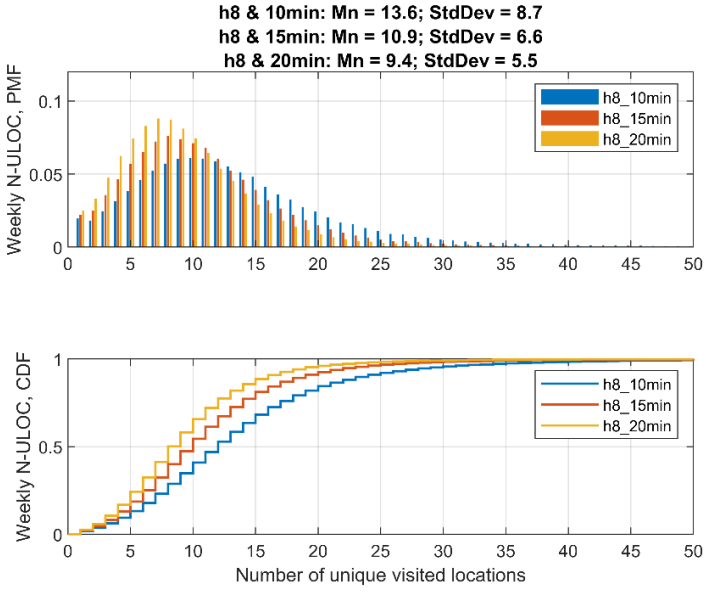

Figure 4-10: Estimated PMF and CDF for N_ULOC weekly: Res. h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 1 to October 7).

Monthly Analysis

The monthly average mean and standard deviation for the different sampling thresholds are shown in Table 4-2. It may be observed that for a bin resolution of h7 and a stay duration threshold of 10 minutes, an individual in the Houston MSA area visits an average of 25.4 unique locations per month with an average standard deviation of around 16.8. Similarly, at longer time scales, increasing the duration threshold while keeping the same bin size results in a lower average number of unique visited locations. Likewise, increasing the bin resolution results in a higher average number of visited locations. Figure 4-11 shows estimated PMFs and CDFs for different spatial sampling thresholds and the say duration threshold of 15 min. It is seen that the curves obtained for the different sampling thresholds are quite similar. Additionally, it appears that N_ULOC is always smaller than 120 on a monthly time scale.

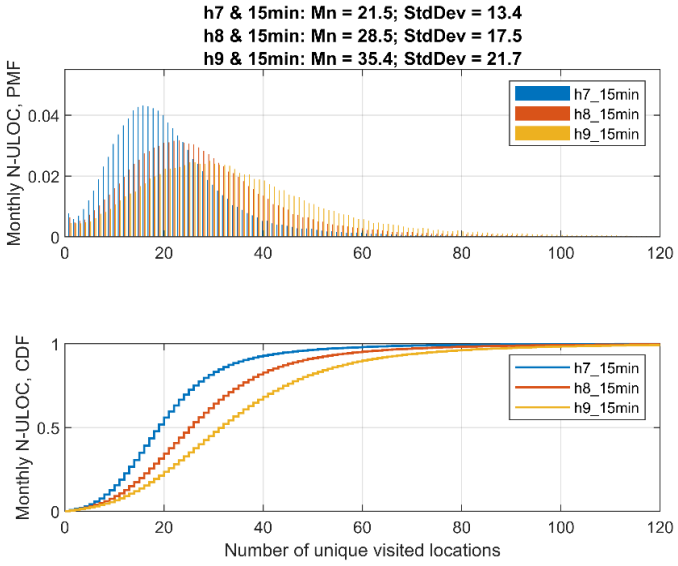

Figure 4-11: Estimated PMF and CDF for N_ULOC monthly: resolution h7, h8, h9 for a stay duration threshold of 15 min (October 2020).

C) Radius of Gyration (R_GYR)

Daily Analysis

For the analysis of R_GYR, the stay duration threshold is no longer relevant. The daily mean and standard deviation for the different sampling thresholds are shown in Table 4-3. It is observed that, for a typical weekday, an individual in the Houston MSA area has an average R_GYR of 12.3 km for resolution h7, 12.2 km for resolution h8, and 12.1 km for resolution h9. The standard deviation is around 13.8 km for all three resolutions. On weekends, the average R_GYR is around 14 km, with a standard deviation of 16.6 km. One can readily notice that, unlike N_LOC and N_ULOC, changing the resolution level for the bin size has a negligible effect on the results. This is understandable since R_GYR measures the deviation of an individual from the centroid of the visited locations. Therefore, the size of the bin should not significantly impact the results. Figure 4-12 shows the average R_GYR on the daily scale for the entire month. It is seen that, unlike N_LOC and N_ULOC, the average measured radius of gyration is higher on weekends, with Saturday being the highest. On the other hand, the lowest measured R_GYR appears to be on Mondays.

An example PMF and CDF set for the R_GYR on a weekday is shown in Figure 4-13. The estimated PMF curves obtained from the three resolution levels are quite similar. It can also be noticed that there is a significant probability of zero R_GYR. This corresponds to the percentage of individuals who stayed at the same location for the whole day. This number is around 16% during weekdays and around 20% during weekends.

Table 4-3: Mean and Standard Deviation for Different Sampling Thresholds R_GYR

	Time Scales	h7	h8	h9
Avg Mean (km)	Weekday	12.3	12.2	12.1
	Weekend	14.0	13.9	13.9
	Week	31.4	31.3	31.2
	Month	51.0	51.0	51.0
Avg Std (km)	Weekday	13.8	13.8	13.8
	Weekend	16.6	16.6	16.6
	Week	23.6	23.6	23.6
	Month	29.3	29.3	29.3

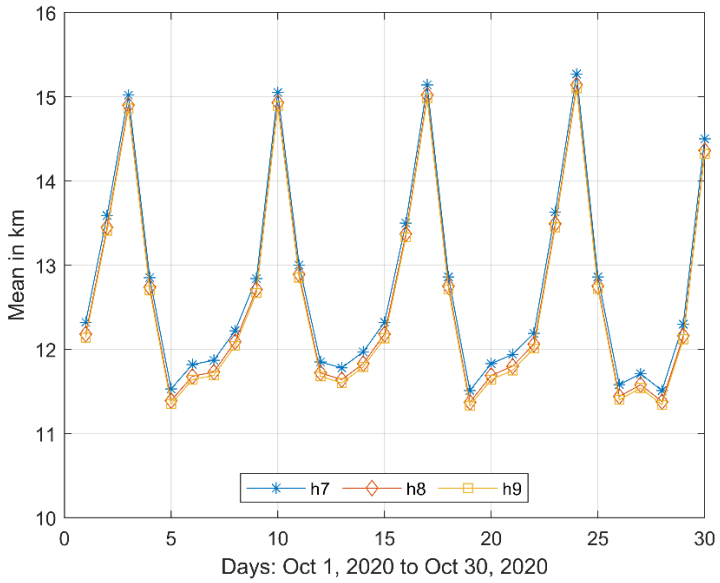

Figure 4-12: Mean of R_GYR on the daily scale.

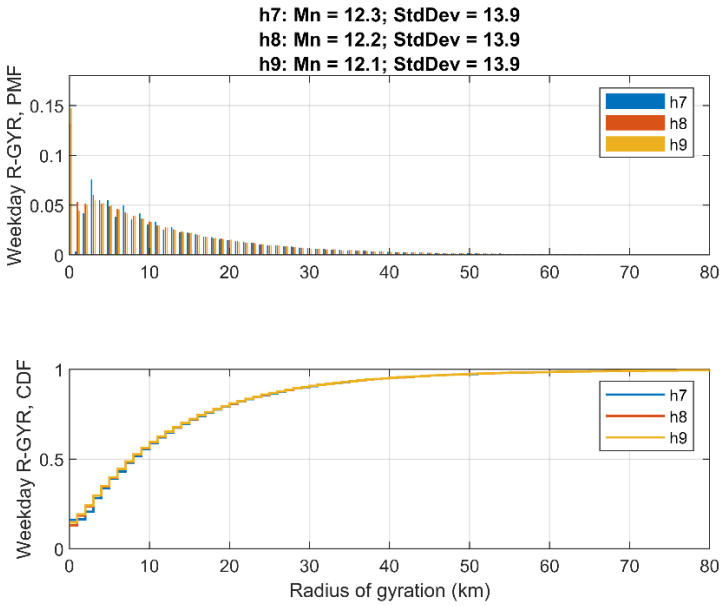

Figure 4-13: Estimated PMF and CDF for R_GYR weekday: resolution h7, h8, and h9 (10/01/2020).

Weekly Analysis

The weekly average and standard deviation for R_GYR at different sampling thresholds are shown in Table 4-3. It is observed that for a typical week, an individual in the Houston MSA area has an average radius of gyration of 31.4 km for resolution h7, 31.3 km for resolution h8, and 31.2 km for resolution h9. The standard deviation is around 23.6 km for all three resolutions. Similar to the daily analysis, changing the resolution level for the bin size has a negligible effect on the results. From Figure 4-14, the estimated PMF curves obtained from the three resolution levels are quite similar. Also, it may be observed that the percentage of individuals with zero R_GYR during the week is around 2%, which is significantly lower compared to the daily time scale.

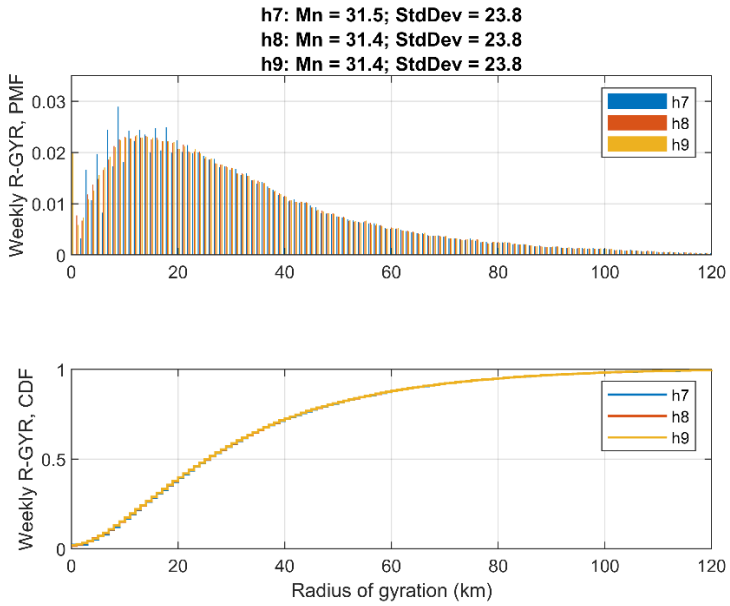

Figure 4-14: Estimated PMF and CDF for R_GYR for a weekly scale: resolution h7, h8, and h9 (October 1 to October 7).

Monthly Analysis

The monthly average mean and standard deviation for the three resolution level thresholds are shown in Table 4-3. It may be observed that on a monthly scale, an individual in the Houston MSA area has an average radius of gyration of 51 km and a standard deviation of 29.3 km for all three spatial resolution levels. Figure 4-15 shows the estimated PMF and CDF for the entire month at different spatial sampling resolutions. The estimated PMF curves obtained from the three resolution levels are quite similar. Also, on the monthly scale, the percentage of individuals with zero R_GYR is only 0.6%.

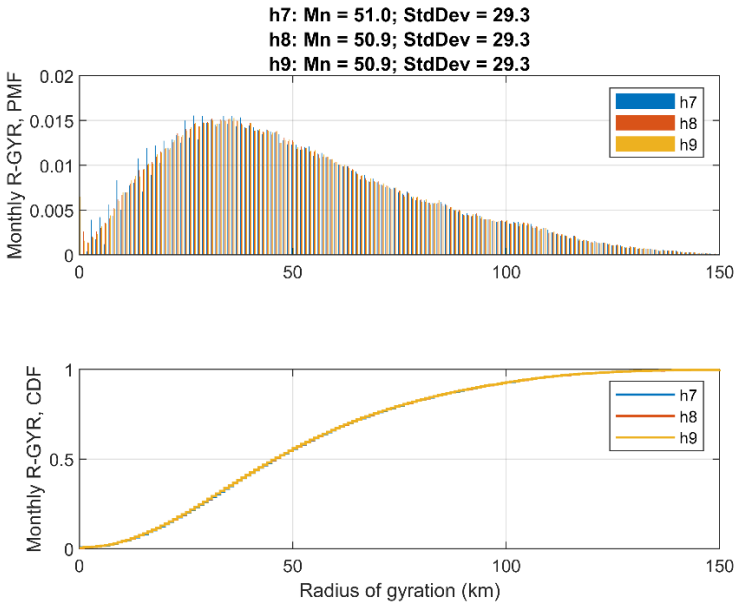

Figure 4-15: Estimated PMF and CDF for R_GYR monthly: resolution h7, h8, and h9 (October 2020).

D) Distance Traveled (D_TRAV)

Daily Analysis

The daily average and standard deviation for the different sampling thresholds are shown in Table 4-4. It is observed that an individual in the Houston MSA area travels an average distance of about 62 km on a weekday and about 70 km on weekends. The results show that there is a slight difference in D_TRAV among the three resolution levels for both weekdays and weekends. This is mainly related to the size of the hexagonal bin.

Figure 4-16 shows the average D_TRAV on the daily scale for the whole month. One notices a distinct seven-day periodicity to the curves. Also, Figure 4-17 shows an example PDF/CDF sets for a sample workday. It is notable that on a daily scale, the D_TRAV almost never exceeds 300 km.

Table 4-4: Mean and Standard Deviation for Different Sampling Thresholds D_TRAV

	Time Scales	h7	h8	h9
Avg Mean (km)	Weekday	67.6	63.2	60.4
	Weekend	66.3	61.9	59.3
	Week	471.2	439.6	420.6
	Month	2035.5	1898.5	1816.0
Avg Std (km)	Weekday	74.5	69.1	66.6
	Weekend	73.5	68.1	65.7
	Week	388.9	359.0	345.3
	Month	1586.0	1461.2	1403.7

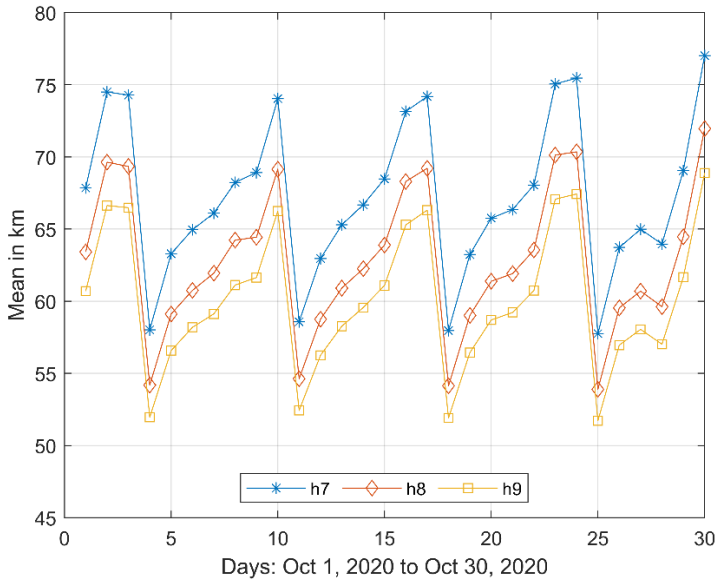

Figure 4-16: Mean of D_TRAV on the daily scale.

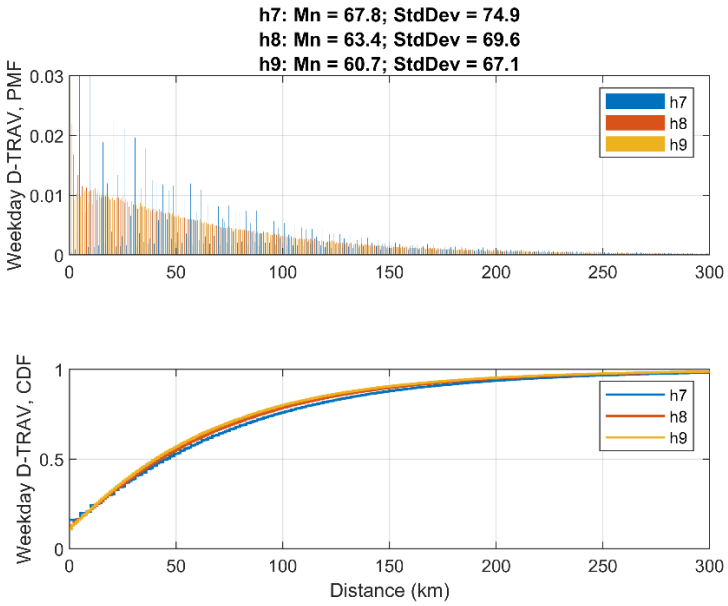

Figure 4-17: Estimated PMF and CDF for D_TRAV weekday: resolution h7, h8, and h9 (10/01/2020).

Weekly Analysis

The weekly average and standard deviation for the different sampling thresholds are presented in Table 4-4. During a typical week, an individual in the Houston MSA area travels an average distance of about 450 km with a standard deviation of about 350 km. The variations due to the sampling resolution are quite small. A sample set of PDFs/CDFs for a weekly scale are shown in Figure 4-18.

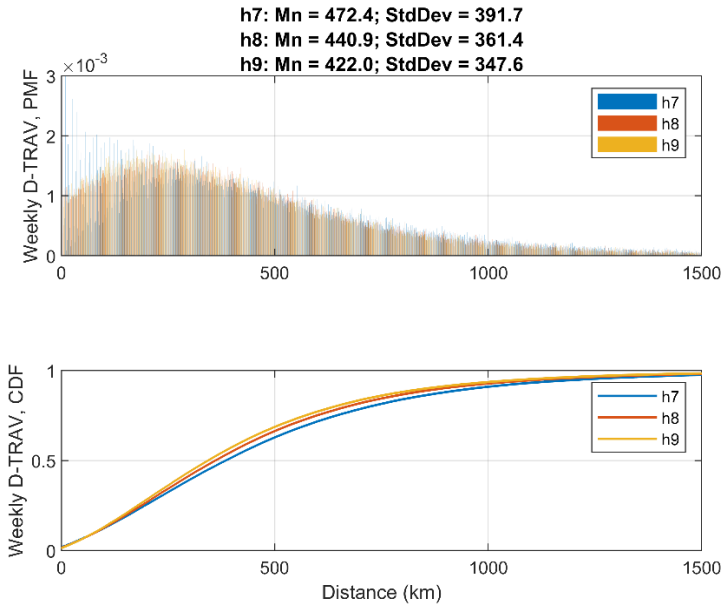

Figure 4-18: Estimated PMF and CDF for D_TRAV weekly: resolution h7, h8, and h9 (October 1 to October 7).

Monthly Analysis

The monthly average and standard deviation for the three sampling thresholds are presented in Table 4-4. On a monthly scale, an individual in the Houston MSA area travels about 1900 km, with a standard deviation of about 1500 km. The values vary within 20% for different spatial resolutions. A set of PDF/CDFs for monthly time scale is presented in Figure 4-19. The curves are quite similar for different spatial resolutions.

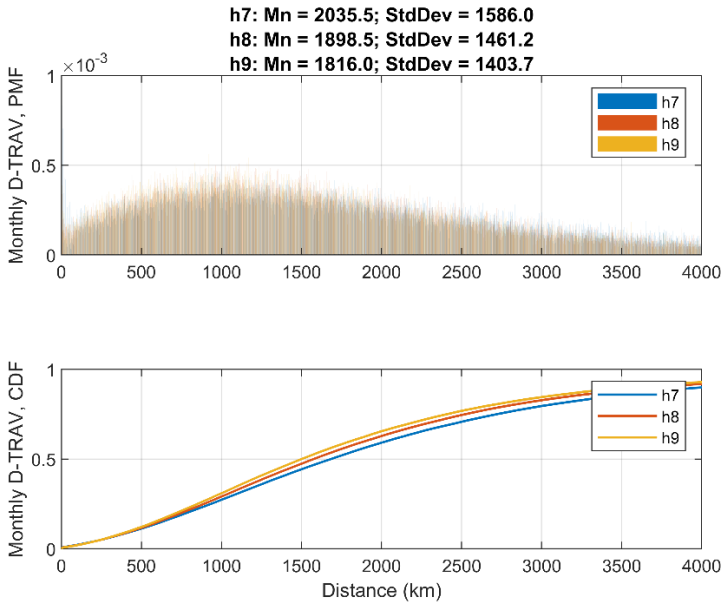

Figure 4-19: Estimated PMF and CDF for D_TRAV monthly: resolution h7, h8, and h9 (October 2020).

New York-Newark-Jersey City MSA

After applying the data preprocessing steps outlined in Chapter 3, the analysis is conducted for each geographical binning sampling threshold (h7, h8, and h9) across daily, weekly, and monthly scales. For daily examinations, weekdays and weekends receive separate attention. For mobility parameters N_LOC, N_ULOC, and T_TP, three stay-duration thresholds of 10, 15, and 20 minutes are examined to identify visited locations at each selected H3 resolution. Records are then filtered to exclude individuals with a "**stay-duration**" shorter than these thresholds, thereby eliminating transient locations with minimal stays. After applying these sampling thresholds, the resultant counts of distinct IDs and locations are summarized in Table 4-5. In the context of this study, each unique bin is regarded as a distinct location.

Table 4-5: Number of IDs and Locations Per Res. Level. (NY-NJ-PA MSA)

	h7	h8	h9
IDs	278742	278742	278742
Locations	4849	31511	190827

Daily Analysis

This analysis focuses on the daily time frame results for the six mobility parameters defined in Chapter 4-Definition. Table 4-6 and Figure 4-20. focus on the Number of Visited Locations (N_LOC) and reveal two main trends. First, an increase in the stay-duration threshold leads to a decrease in the average number of visited locations. This decrease occurs because higher stay-duration thresholds mean that fewer locations meet the criteria to be considered "visited". The locations that are omitted are typically transient or simply pass-through locations. Second, employing a finer hexagonal bin resolution level results in an increase in the average number of visited locations, as a more detailed geographic grid captures a greater number of unique places.

Figure 4-20 reinforces these findings by introducing a temporal dimension; it shows that people generally visit more locations on weekends than weekdays. However, an exception arises when a 20-minute stay-duration threshold is combined with an h9 bin resolution.

Figures 4-21 to 4-24 provide additional insights, presenting the Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) for N_LOC. These figures indicate that people predominantly stay in one location, most often their home, particularly when analyzed across different sampling thresholds. As the number of visits increases, the PMF curve generally declines. Notably, exceptions exist; for example, a dip in the number of visits to two locations can be observed in

the figures, particularly when people travel to unique locations like airports and not returning that day. Importantly, the stability of the PMF curves across these thresholds confirms the robustness of the N_LOC parameter.

Turning to the Number of Unique Visited Locations (N_ULOC), Table 4-7 identifies two main trends, which Figure 4-25 corroborates. First, just like with N_LOC, a higher stay-duration threshold reduces the average number of unique locations visited. Second, a more detailed hexagonal bin resolution increases this number. Figure 4-25 further reveals that fewer unique locations are visited on weekdays compared to weekends, with Fridays being the most mobile days and Sundays the least.

For an extended understanding, Figures 4-26 to 4-29 explore N_ULOC's PMF and CDF at multiple stay-duration thresholds across various bin resolutions. They align well with observations made from Table 4-7 and Figure 4-25 and also demonstrate the metric's stability across different sampling conditions, similar to N_LOC.

Figure 4-28 and Figure 4-29 specifically show the impact of varying stay-duration thresholds on N_ULOC at h8 bin resolution, again confirming the metric's reliability.

In terms of N_SIG, Table 4-8 and Figure 4-30 offer insights into this daily mobility parameter across various hexagonal bin resolutions. Here, the trend is incremental: the average number of significant locations visited slightly increases when bin resolution goes from h7 to h9. People tend to visit more significant locations on weekdays than weekends, a deviation from trends observed in N_LOC and N_ULOC. Nevertheless, Fridays remain peak days for mobility in this parameter as well, while Sundays record the least activity. The modest fluctuation between

weekdays and weekends suggests that most people's daily routines generally involve one or two key locations, like home and work.

Figure 4-31 and Figure 4-32 further corroborate the data's consistency, revealing that over 60% of people visited just one significant location across all resolutions.

For displacement behavior parameters, Table 4-9 and Figure 4-33 focus on the radius of gyration (R_GYR). Unlike N_LOC and N_ULOC, R_GYR shows considerable stability across bin resolutions. Figure 4-33 shows a peak in R_GYR values during weekends, particularly on Saturdays, suggesting increased spatial mobility.

Figure 4-34 and Figure 4-35 present the PMF and CDF curves for R_GYR, revealing regularity but also minor discrepancies in the h7 curve for values below 5 km.

The dataset also shows an increase in zero R_GYR values from weekdays to weekends, indicating a potential preference among people to limit their spatial activities during the weekend.

Table 4-10 and Figure 4-36 explore the D_TRAV (distance traveled) parameter, revealing minor discrepancies in average distances based on bin resolution. This slight variation is influenced by hexagonal bin size, which affects distance calculations based on center coordinates of binned locations. In contrast, R_GYR, measuring maximum distance from the mobility center of mass, shows less sensitivity to bin size. Hence, higher binning resolutions are advised for more accurate distance calculations.

Figure 4-37 and Figure 4-38 provide the PMF and CDF curves for D_TRAV, confirming its consistency regardless of bin resolution or day of the week.

Finally, Table 4-11 and Figure 4-39 summarize the behavior of T_TP parameter. A higher bin resolution corresponds to a higher average T_TP, and stay duration also has a clear impact, with T_TP values rising consistently as the duration threshold increases. Figure 4-39 also shows higher travel time percentages during weekends, especially Saturdays, compared to weekdays, particularly Mondays.

The PMF curves presented in Figures 4-40 to 4-43 reveal consistent patterns across sampling thresholds, confirming the metric's stability and consistency across different sampling thresholds.

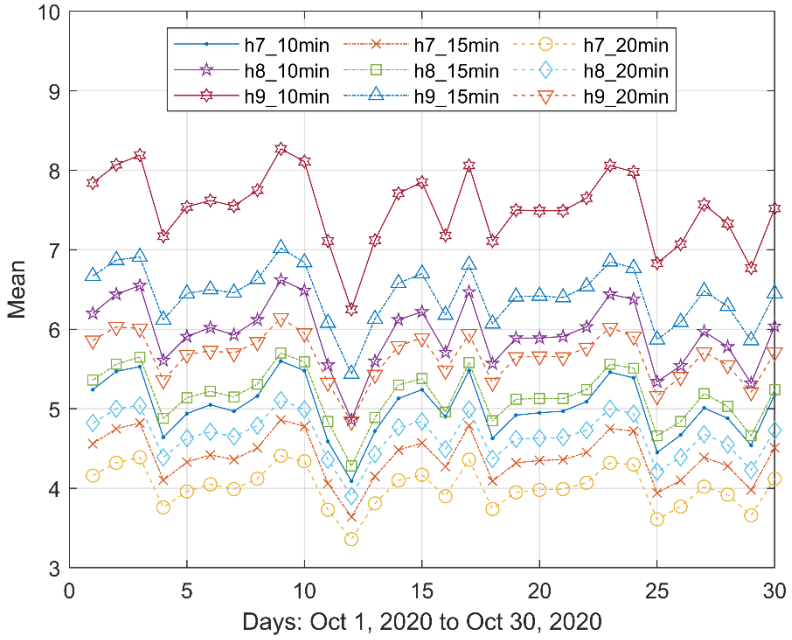

Figure 4-20: Mean of N_LOC between 10/01/2020 and 10/30/2020

Table 4-6: N_LOC: Average Mean and Standard Deviation for the Different Sampling Thresholds

	Time Scales	h7			h8			h9		
		10 Min	15 Min	20 Min	10 Min	15 Min	20 Min	10 Min	15 Min	20 Min
Avg Mean	Weekday	5.01	4.38	4.01	5.93	5.15	4.65	7.51	6.43	5.67
	Weekend	5.02	4.41	4.03	5.99	5.20	4.66	7.57	6.43	5.62
	Week	29.53	25.19	22.57	36.32	30.83	27.27	47.68	40	34.61
	Month	123.05	104.53	93.34	151.99	128.59	113.44	200.61	167.93	145.03
Avg. StdDev	Weekday	4.23	3.38	2.87	4.95	4.03	3.41	6.27	5.15	4.30
	Weekend	4.08	3.30	2.82	4.87	4.00	3.39	6.26	5.16	4.29
	Week	23.55	18.85	15.96	28.43	23.42	19.84	37.32	31.17	26.12
	Month	96.4	77.15	65.25	116.12	95.87	81.28	152.61	128.05	107.5

Table 4-7: N_ULOC: Average Mean and Standard Deviation for the Different Sampling Thresholds

	Time Scales	h7			h8			h9		
		10 Min	15 Min	20 Min	10 Min	15 Min	20 Min	10 Min	15 Min	20 Min
Avg Mean	Weekday	2.98	2.6	2.39	3.37	2.89	2.61	3.87	3.25	2.87
	Weekend	3.06	2.68	2.46	3.49	2.99	2.69	4.02	3.35	2.93
	Week	9.39	7.92	7.04	12.1	9.85	8.49	15.05	11.8	9.81
	Month	21.27	17.95	15.99	30.53	24.74	21.26	41.28	31.9	26.2
Avg. StdDev	Weekday	2.12	1.64	1.39	2.34	1.8	1.52	2.67	2.06	1.71
	Weekend	2.06	1.61	1.38	2.29	1.79	1.53	2.67	2.07	1.71
	Week	6.5	5.06	4.27	8.07	6.03	5.01	9.68	7.14	5.75
	Month	14.66	11.66	9.99	21.1	15.67	12.99	27.38	19.76	15.82

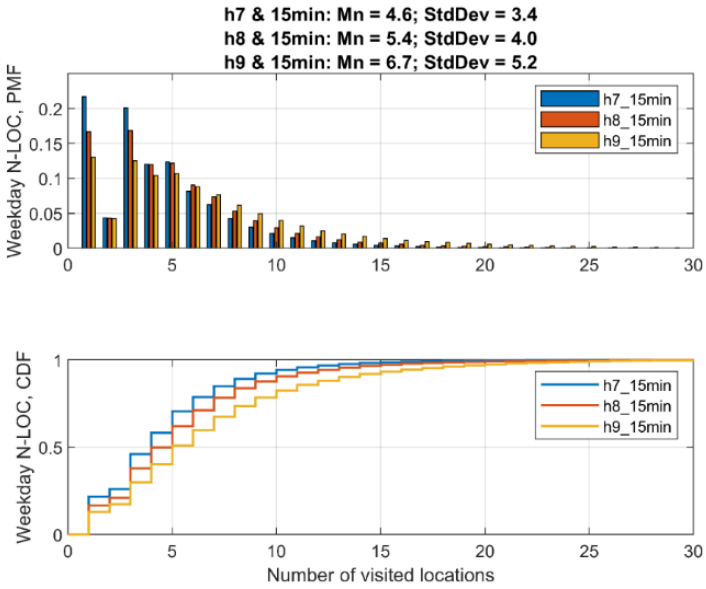

Figure 4-21: Estimated PMF and CDF for N_LOC weekday: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/01/2020).

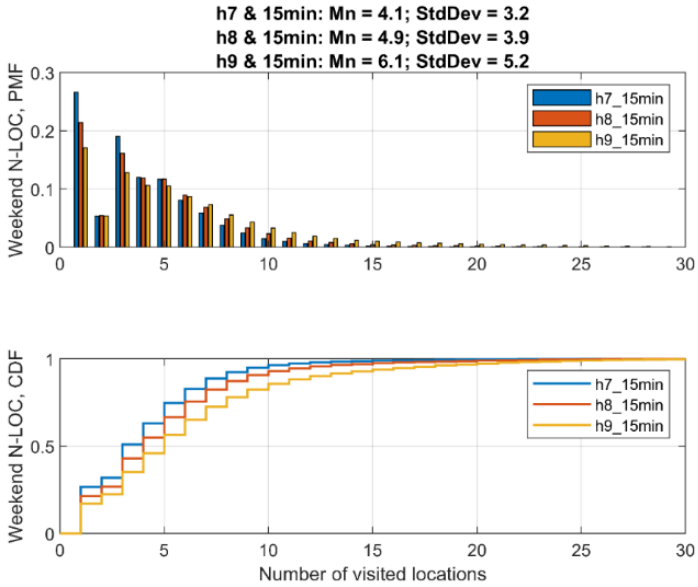

Figure 4-22: Estimated PMF and CDF for N_LOC weekend: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/04/2020).

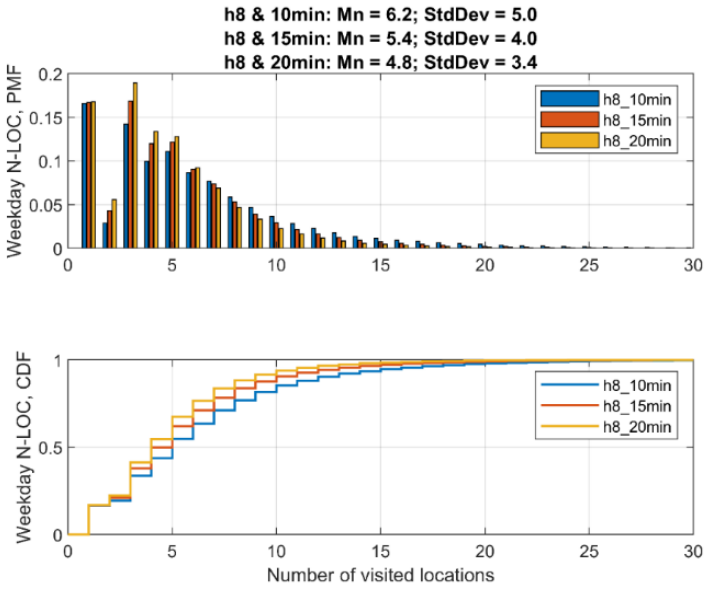

Figure 4-23: Estimated PMF and CDF for N_LOC weekday: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/01/2020).

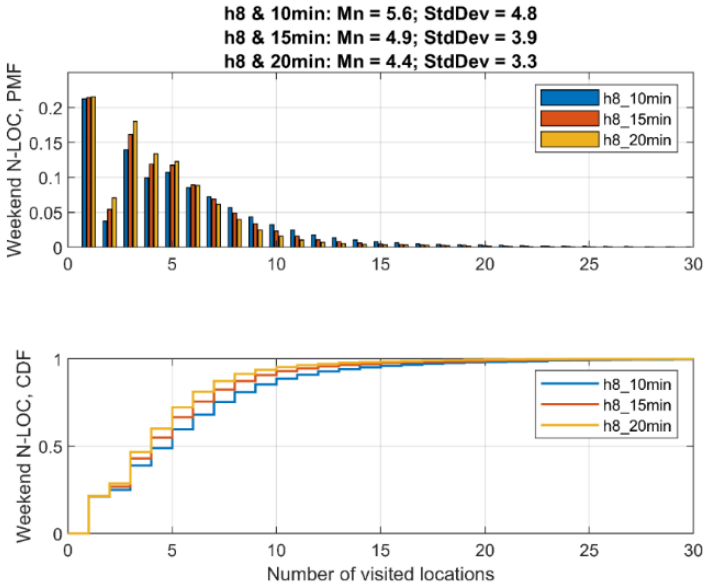

Figure 4-24: Estimated PMF and CDF for N_LOC weekend: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/04/2020).

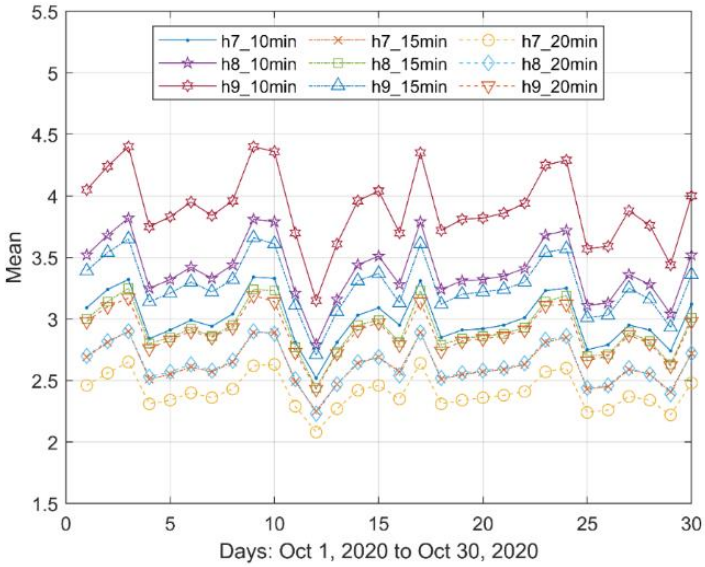

Figure 4-25: Mean of N_ULOC for the period between 10/01/2020 and 10/30/2020.

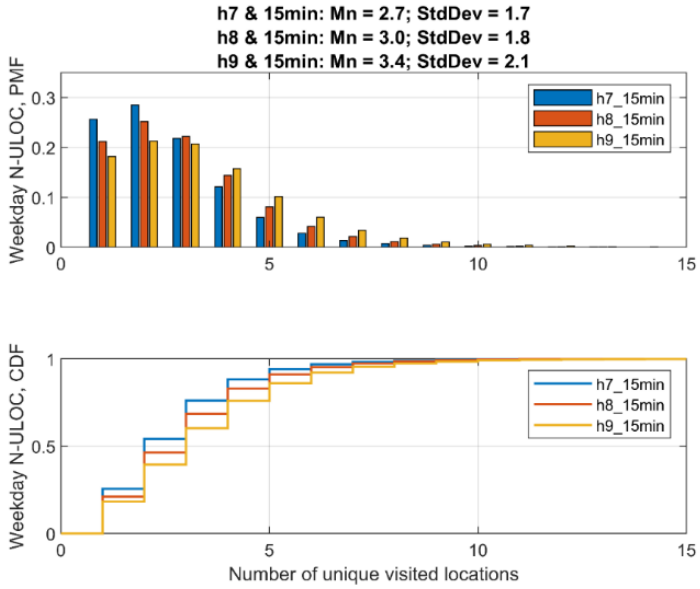

Figure 4-26: Estimated PMF and CDF for N_ULOC weekday: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/01/2020).

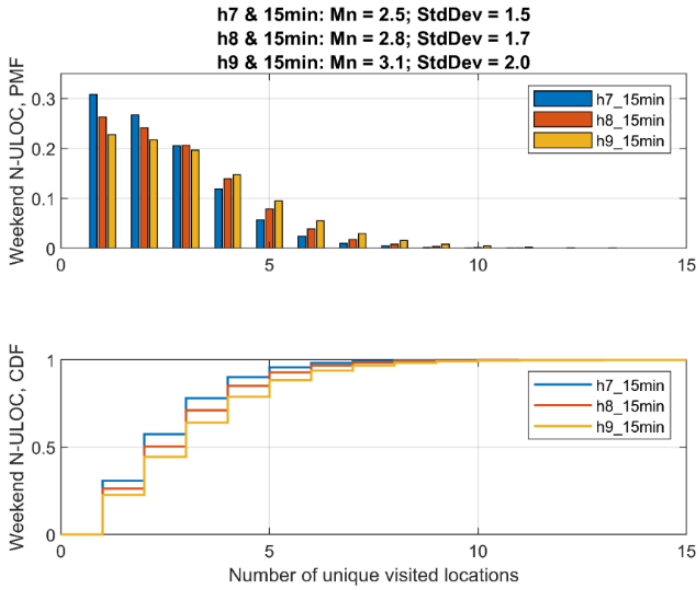

Figure 4-27: Estimated PMF and CDF for N_ULOC weekend: resolution h7, h8, and h9 for a stay duration threshold of 15 min (10/04/2020).

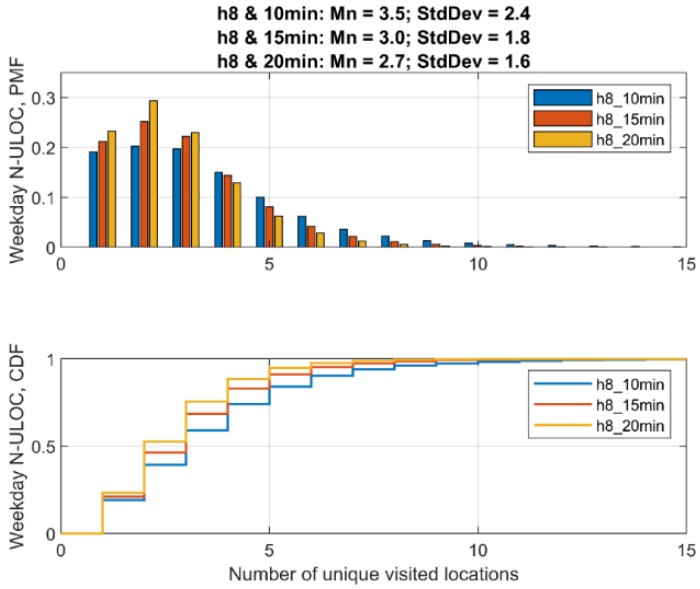

Figure 4-28: Estimated PMF and CDF for N_ULOC weekday: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/01/2020).

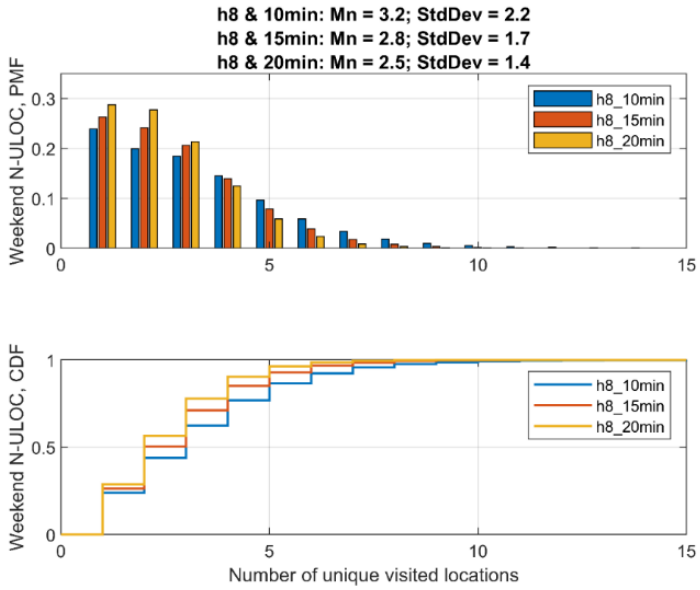

Figure 4-29: Estimated PMF and CDF for N_ULOC weekend: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/04/2020).

Table 4-8: N_SIG: Average Mean and Std for Different Sampling Thresholds.

	Time Scales	h7	h8	h9
Avg Mean (km)	**Weekday**	1.38	1.39	1.4
	Weekend	1.24	1.25	1.26
	Week	1.3	1.31	1.31
	Month	1.29	1.29	1.29
Avg Std (km)	**Weekday**	0.51	0.52	0.53
	Weekend	0.45	0.46	0.47
	Week	0.48	0.49	0.5
	Month	0.47	0.48	0.48

Table 4-9: R_GYR: Average Mean and Std for Different Sampling Thresholds.

	Time Scales	h7	h8	h9
Avg Mean (km)	**Weekday**	8.78	8.65	8.61
	Weekend	11.57	11.45	11.41
	Week	26.47	26.33	26.29
	Month	46.45	46.32	46.29
Avg Std (km)	**Weekday**	12.03	12.01	12.01
	Weekend	16.48	16.48	16.48
	Week	24.83	24.84	24.85
	Month	33.95	33.97	33.98

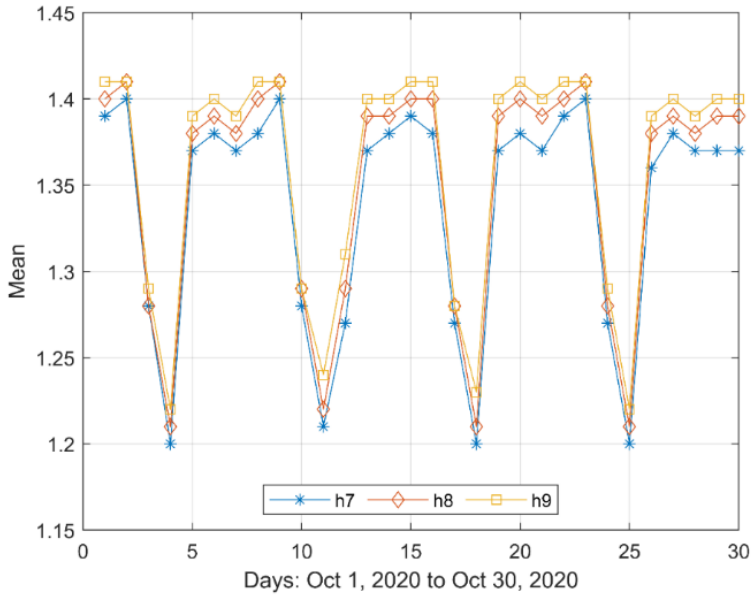

Figure 4-30: Mean of N_SIG between 10/01/2020 and 10/30/2020.

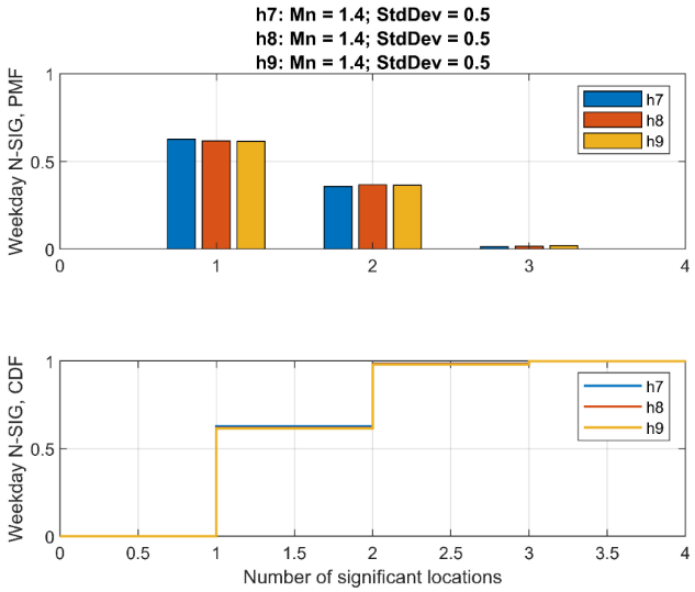

Figure 4-31: Estimated PMF and CDF for N_SIG weekday: resolution h7, h8, and h9 (10/01/2020).

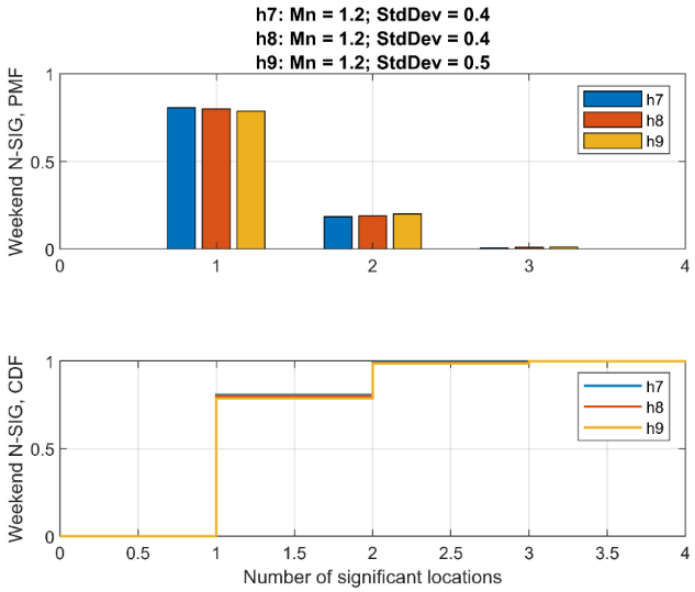

Figure 4-32: Estimated PMF and CDF for N_SIG weekend: resolution h7, h8, and h9 (10/04/2020).

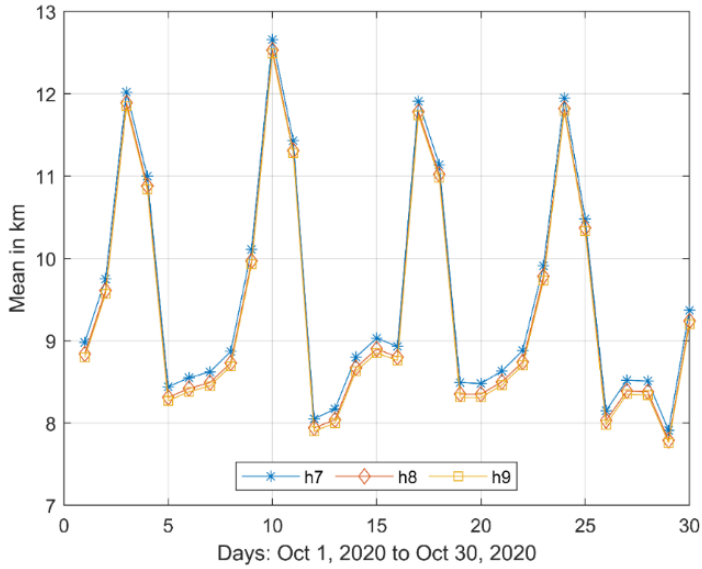

Figure 4-33: Mean of R_GYR between 10/01/2020 and 10/30/2020.

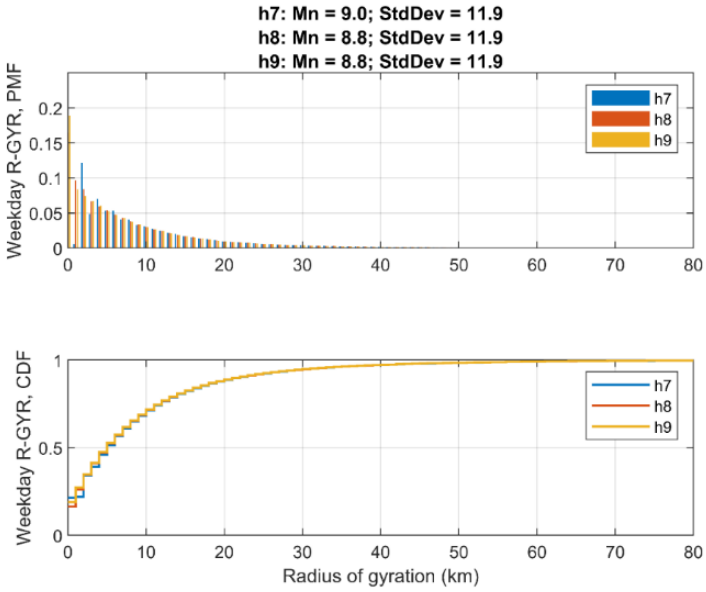

Figure 4-34: Estimated PMF and CDF for R_GYR weekday: resolution h7, h8, and h9 (10/01/2020).

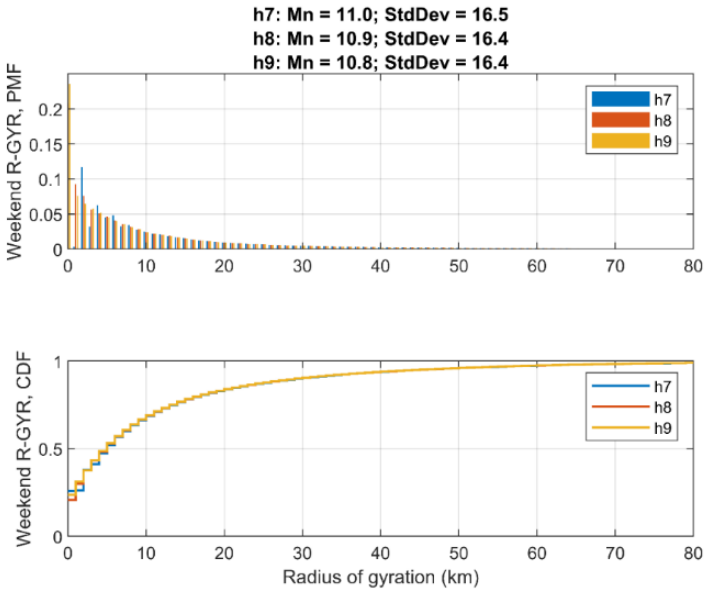

Figure 4-35: Estimated PMF and CDF for R_GYR weekend: resolution h7, h8, and h9 (10/04/2020).

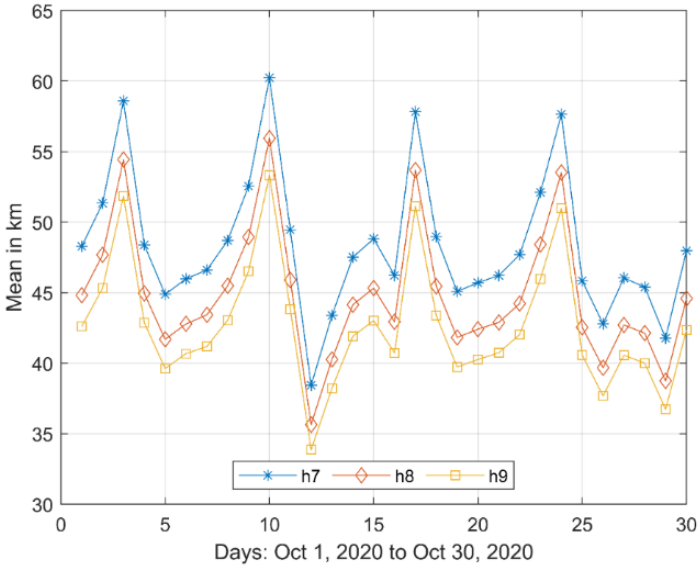

Figure 4-36: Mean of D_TRAV between 10/01/2020 and 10/30/2020.

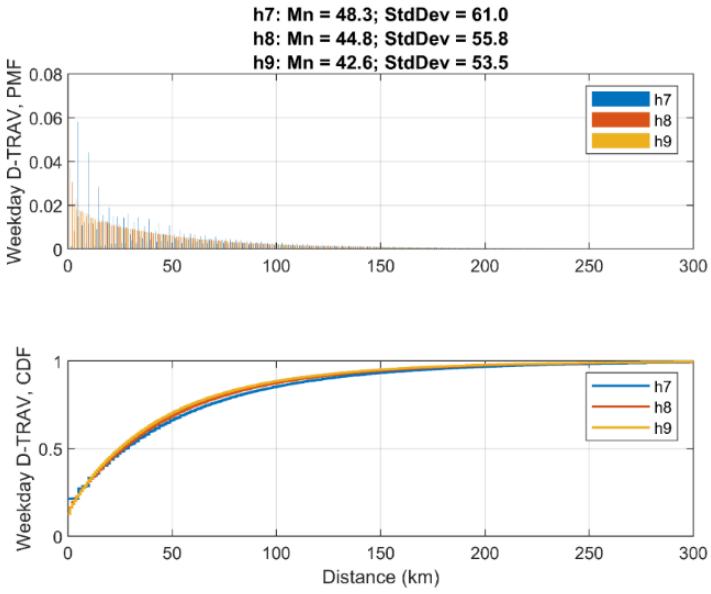

Figure 4-37: Estimated PMF and CDF for D_TRAV weekday: resolution h7, h8, and h9 (10/01/2020).

Table 4-10: D_TRAV: Average Mean and Standard Deviation for Different Sampling Thresholds.

	Time Scales	h7	h8	h9
Avg Mean (km)	**Weekday**	46.52	43.22	41.04
	Weekend	53.37	49.56	47.24
	Week	341.76	317.2	301.56
	Month	1458.22	1353.33	1286.45
Avg Std (km)	**Weekday**	60.53	55.44	53.02
	Weekend	67.37	61.87	59.5
	Week	320.56	292.83	279.53
	Month	1298.64	1185.45	1130.2

Table 4-11: T_TP: Average Mean and Standard Deviation for the Different Sampling Thresholds.

	Time Scales	h7			h8			h9		
		10 Min	15 Min	20 Min	10 Min	15 Min	20 Min	10 Min	15 Min	20 Min
Avg Mean	Weekday	2.57	3.12	3.59	3.74	4.42	5.06	4.68	5.61	6.6
	Weekend	2.87	3.42	3.91	4.01	4.72	5.4	4.98	5.97	7.03
	Week	3.08	3.6	4.05	4.18	4.83	5.45	5.09	5.99	6.95
	Month	2.66	3.17	3.62	3.76	4.4	5.02	4.66	5.55	6.5
Avg. StdDev	Weekday	3.59	4.3	4.92	4.77	5.53	6.19	5.56	6.43	7.25
	Weekend	4.03	4.65	5.2	5.02	5.7	6.3	5.73	6.54	7.33
	Week	2.78	3.26	3.69	3.61	4.14	4.59	4.16	4.77	5.35
	Month	2.35	2.85	3.3	3.24	3.77	4.21	3.8	4.4	4.96

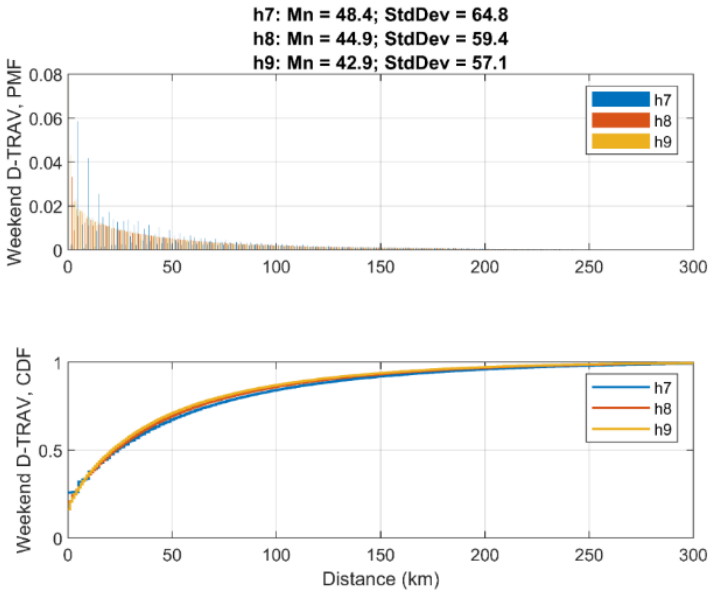

Figure 4-38: D_TRAV PMF and CDF weekend: (10/04/2020).

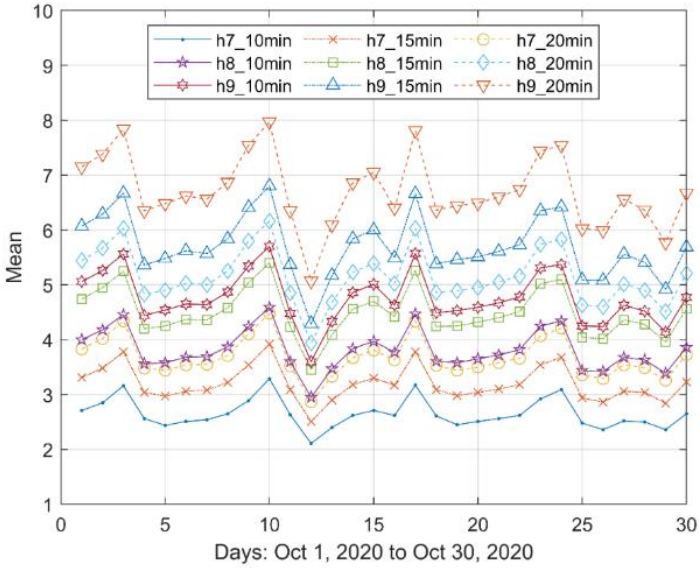

Figure 4-39: Mean of T_TP between 10/01/2020 and 10/30/2020.

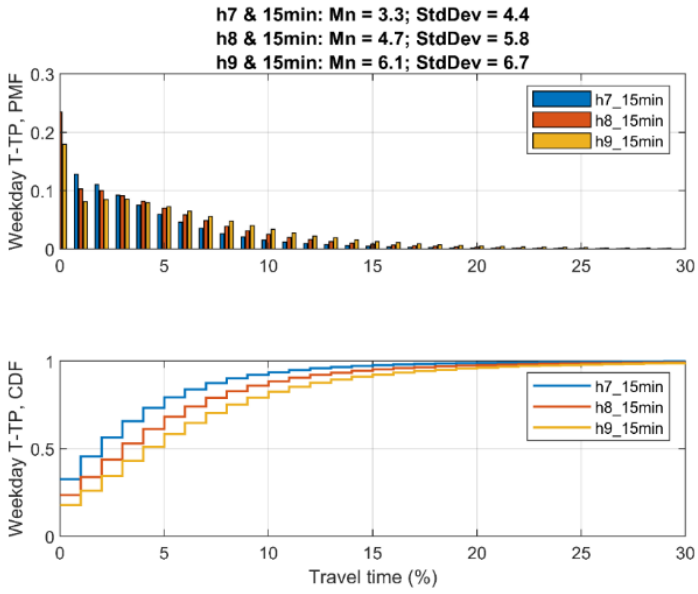

Figure 4-40: Estimated PMF and CDF for T_TP weekday (10/01/2020).

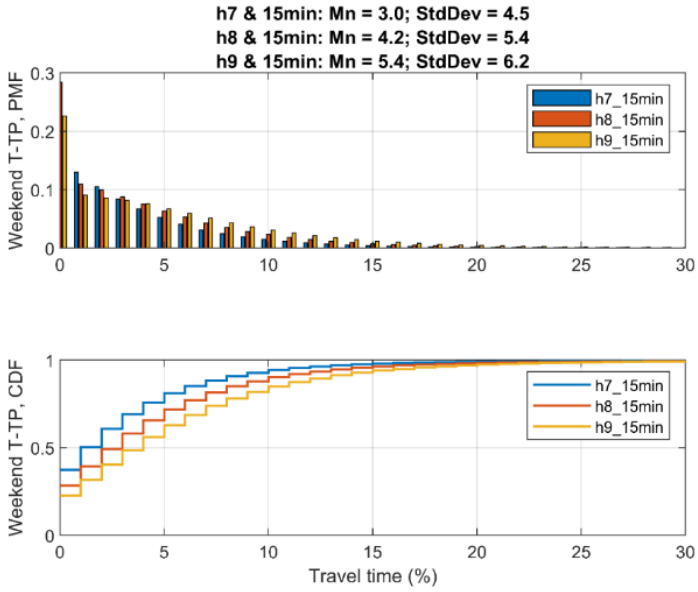

Figure 4-41: Estimated PMF and CDF for T_TP weekend (10/04/2020).

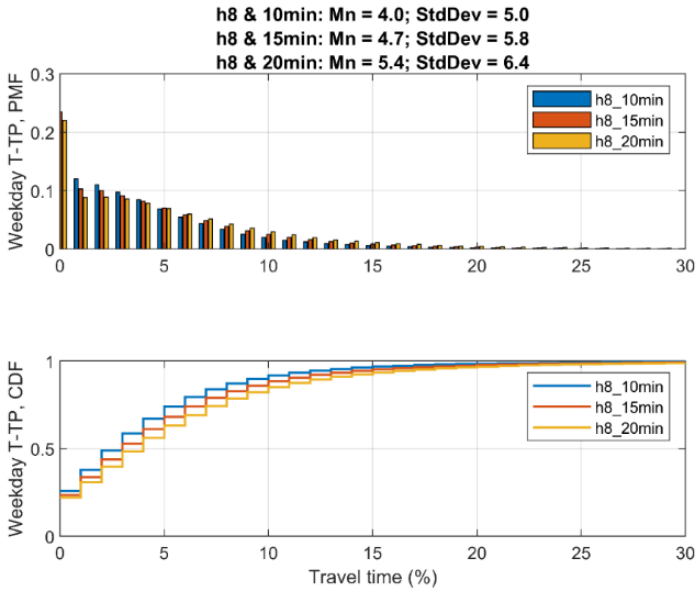

Figure 4-42: Estimated PMF and CDF for T_TP weekday: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/01/2020).

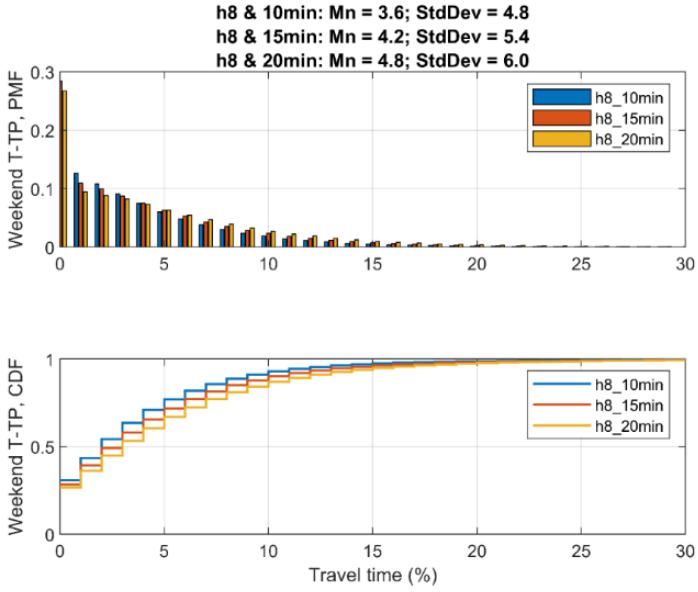

Figure 4-43: Estimated PMF and CDF for T_TP weekend: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (10/04/2020).

Weekly Analysis

The weekly analysis results of the Number of Visited Locations (N_LOC) are presented in Table 4-6 and Table 4-12. Specifically, Table 4-6 shows how an increase in either bin resolution or stay duration threshold elevates the average number of weekly visited locations. While there are daily variations in this parameter, the study confirms that such fluctuations are minimal on a weekly basis as shown in Table 4-12. Notably, N_LOC remains consistently below 180 on a weekly time scale.

Figure 4-44 and Figure 4-45 illustrate the Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) for N_LOC across various resolutions and stay duration thresholds, respectively. These figures reveal a general similarity in patterns, mirroring the daily analysis results, indicating that the data is fairly consistent across different sampling conditions.

Regarding the Number of Unique Visited Locations (N_ULOC), Table 4-7 and Table 4-13 reveal that the average number of unique locations visited weekly increases with higher bin resolution or longer duration thresholds. This pattern aligns with the daily analysis and mirrors the findings on N_LOC. Figure 4-46 and Figure 4-47 further illustrate this trend, displaying PMF and CDF curves that show slight differences but are largely consistent across resolutions and thresholds. Notably, N_ULOC remains consistently below 50 on a weekly time scale.

The Number of Significant Locations (N_SIG) results are presented in Table 4-8 and Table 4-14, offering a slightly different perspective. Changing the bin resolution has only a minor impact on the average number of significant locations visited, a finding that aligns with daily analyses. Figure 4-48 underscores this consistency, displaying PMF curves that are nearly identical across different bin resolutions and thresholds.

Table 4-9 and Table 4-15 focus on the Radius of Gyration (R_GYR) and reveals an intriguing observation: altering the bin resolution has a negligible effect on the average R_GYR. This finding reflects results from daily analyses, where R_GYR remained largely stable. Figure 4-49 supports this observation, displaying PMF and CDF curves that are highly similar, with some variations noticeable for R_GYR values less than 20 km. An important observation is that only about 3% of individuals show a zero R_GYR during the week, a rate significantly lower compared to the daily time scale.

For Distance Traveled (D_TRAV), Table 4-10 and Table 4-16 indicate that changes in bin resolution have a noticeable effect on the average distance covered, aligning with daily fluctuations. Figure 4-50 corroborates these findings, displaying PMF and CDF curves that are generally consistent but show noticeable variations for distances less than 100 km. Analysis of weekly data for resolution h7, as presented in Table 4-16, reveals that the average distance traveled (D_TRAV) shows minimal changes from week to week. This finding contrasts with the variations observed in day-to-day mobility and is consistent with other studied parameters.

Finally, the Travel Time Percentage (T_TP) is presented in Table 4-11 and Table 4-17. These tables show that an increase in bin resolution and stay duration threshold results in a higher average percentage of time spent traveling. This is consistent with daily analyses, and Figure 4-51 and Figure 4-52 confirm this, displaying PMF and CDF curves across different sampling thresholds that are quite similar, albeit with minor variations in the peaks.

Table 4-12: N_LOC: Weekly Mean and Standard Deviation for Resolution h7, 10 Min

Week	Mean	StdDev
Oct. 1 to Oct. 7	30.18	23.73
Oct. 8 to Oct. 14	29.15	23.22
Oct. 15 to Oct. 21	29.45	23.66
Oct. 22 to Oct.28	29.32	23.59

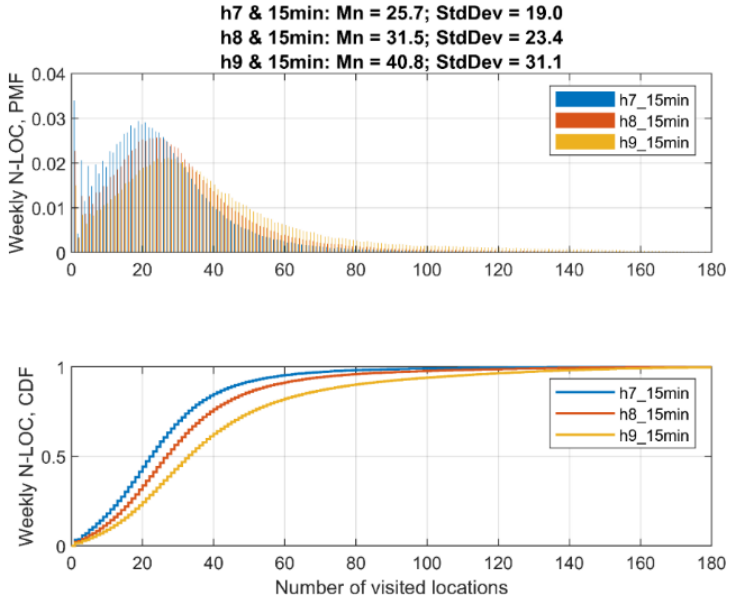

Figure 4-44: Estimated PMF & CDF for N_LOC weekly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 1 - October 7).

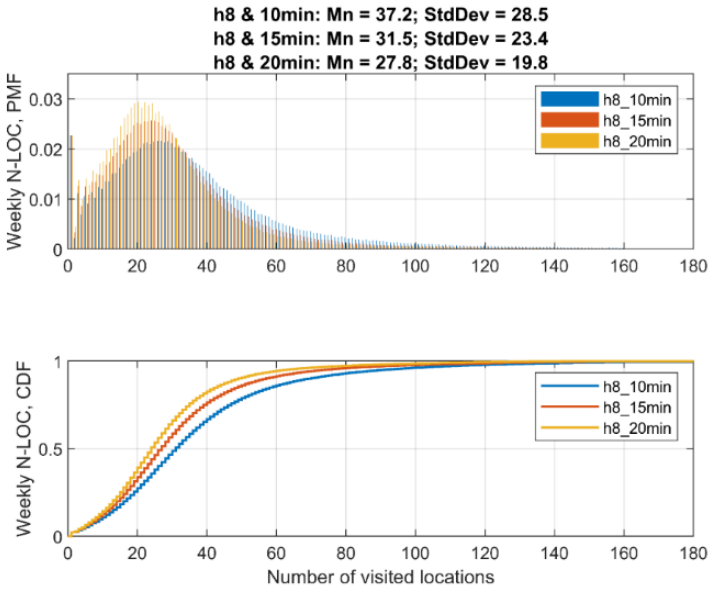

Figure 4-45: Estimated PMF and CDF for N_LOC weekly: Res. h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 1 to October 7).

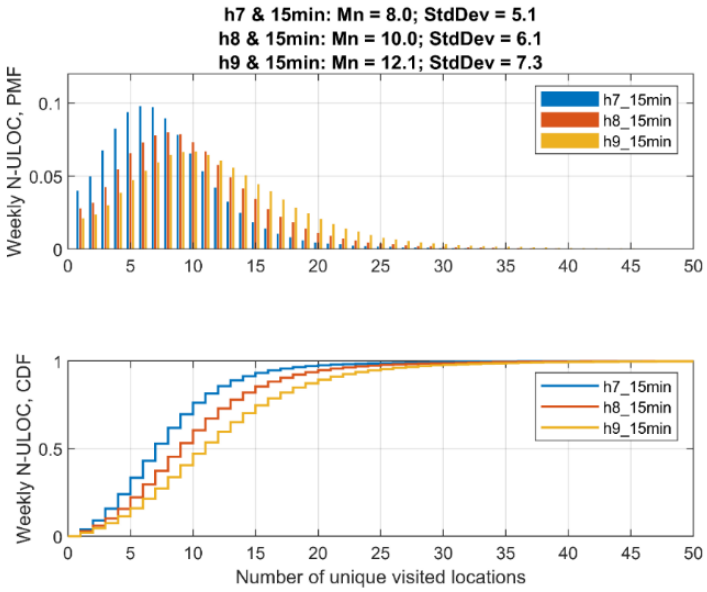

Figure 4-46: Estimated PMF and CDF for N_ULOC weekly: Res. h7, h8, and h9 for a stay duration threshold of 15 min (October 1 to October 7).

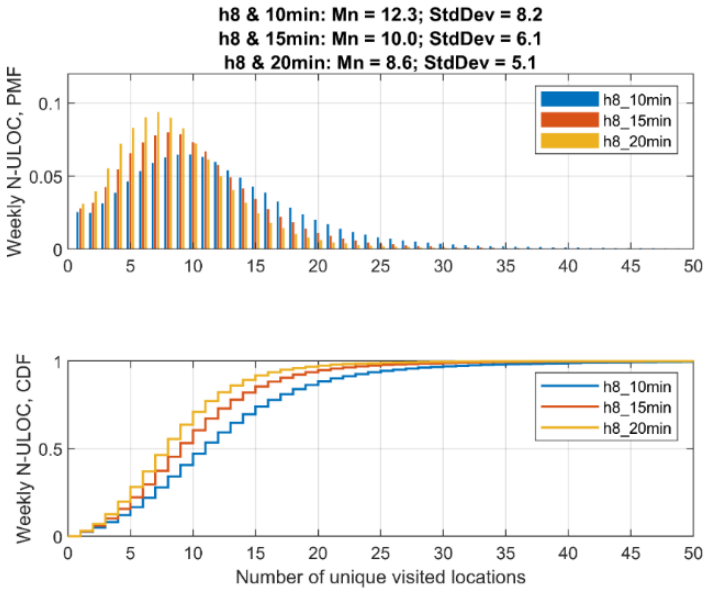

Figure 4-47: Estimated PMF and CDF for N_ULOC weekly: Res.h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 1 to October 7).

Table 4-13: N_ULOC: Weekly Mean and Standard Deviation for Resolution h7, 10 Min

Week	Mean	StdDev
Oct. 1 to Oct. 7	9.53	6.58
Oct. 8 to Oct. 14	9.47	6.47
Oct. 15 to Oct. 21	9.31	6.5
Oct. 22 to Oct.28	9.26	6.44

Table 4-14: N_SIG: Weekly Mean and Standard Deviation for Resolution h7

Week	Mean	StdDev
Oct. 1 to Oct. 7	1.31	0.49
Oct. 8 to Oct. 14	1.29	0.48
Oct. 15 to Oct. 21	1.31	0.48
Oct. 22 to Oct.28	1.3	0.48

Table 4-15: R_GYR: Weekly Mean and Standard Deviation for Resolution h7

Week	Mean	StdDev
Oct. 1 to Oct. 7	26.41	24.69
Oct. 8 to Oct. 14	27.54	25.74
Oct. 15 to Oct. 21	26.11	24.57
Oct. 22 to Oct.28	25.83	24.31

Table 4-16: D_TRAV: Weekly Mean and Standard Deviation for Resolution h7

Week	Mean	StdDev
Oct. 1 to Oct. 7	345.69	322.87
Oct. 8 to Oct. 14	341.92	316.22
Oct. 15 to Oct. 21	340.4	321.91
Oct. 22 to Oct.28	339.02	321.23

Table 4-17: T_TP: Weekly Mean and Standard Deviation for Resolution h7

Week	Mean	StdDev
Oct. 1 to Oct. 7	3.22	2.9
Oct. 8 to Oct. 14	3.03	2.7
Oct. 15 to Oct. 21	3.04	2.76
Oct. 22 to Oct.28	3.04	2.77

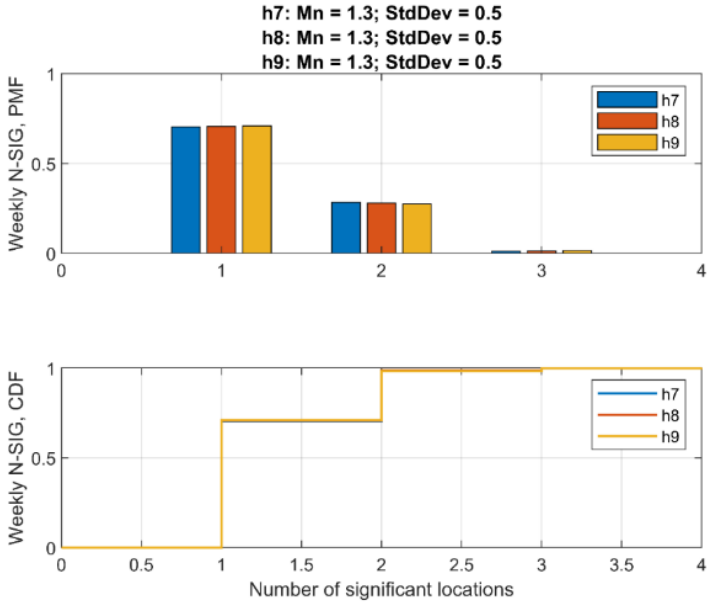

Figure 4-48: Estimated PMF & CDF for N_SIG weekly: (Oct 1 to Oct 7).

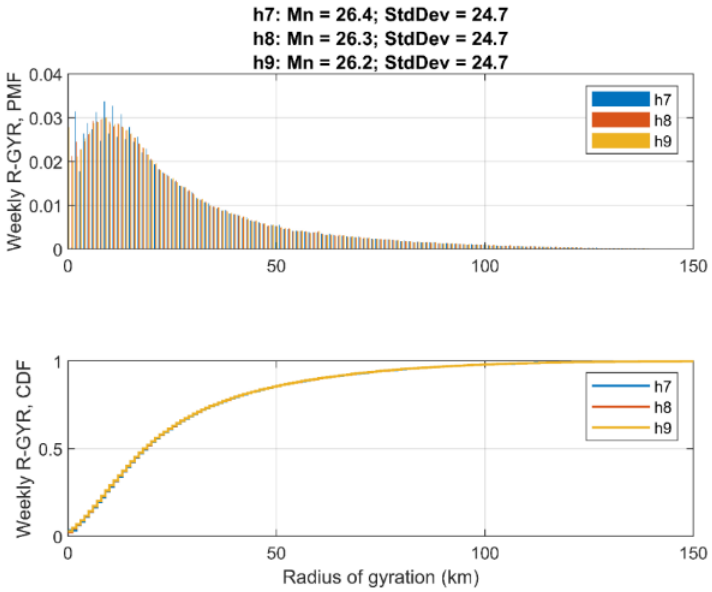

Figure 4-49: Estimated PMF & CDF for R_GYR weekly (Oct 1 to Oct 7).

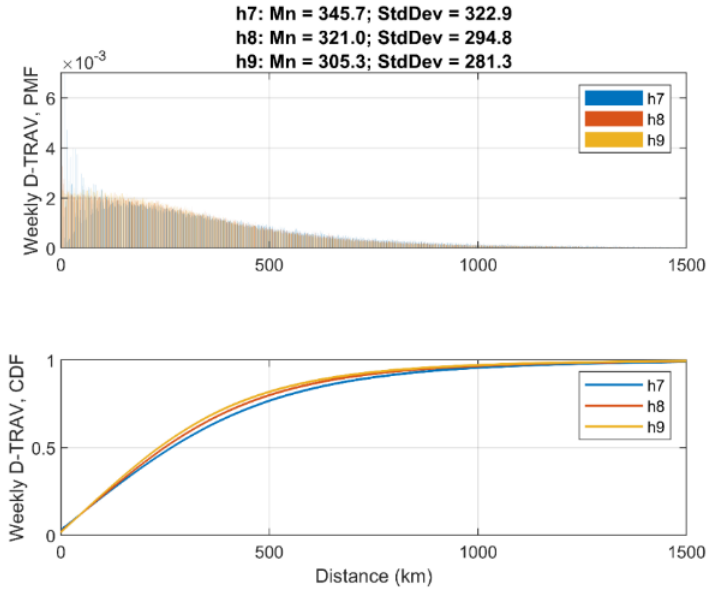

Figure 4-50: Estimated PMF & CDF for D_TRAV weekly (Oct 1 to Oct 7).

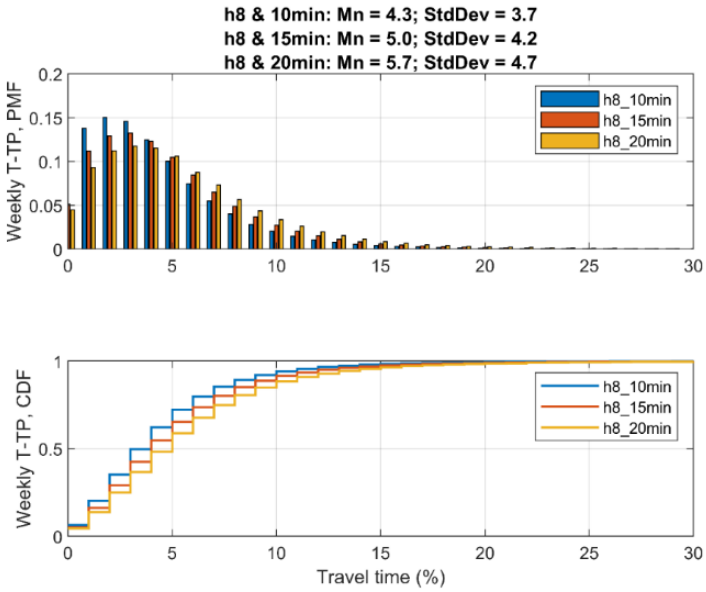

Figure 4-51: Estimated PMF and CDF for T_TP weekly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 1 to October 7).

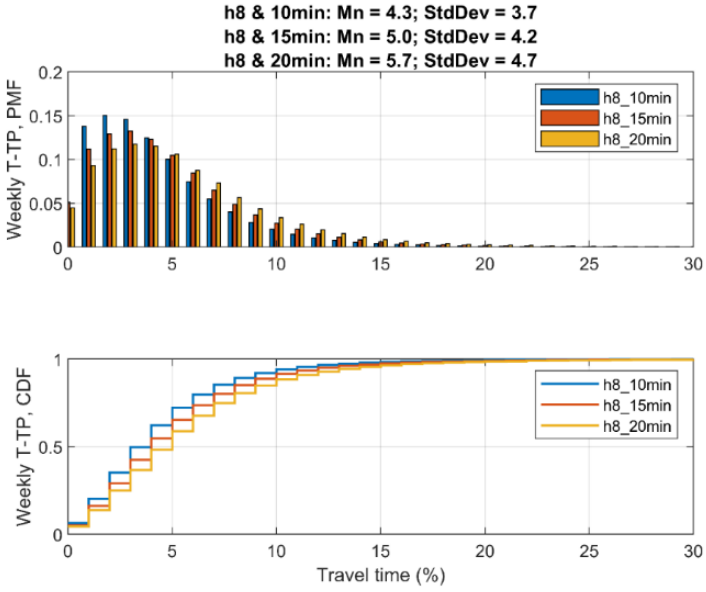

Figure 4-52: Estimated PMF and CDF for T_TP weekly: Res. h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 1 to October 7).

Monthly Analysis

Starting with the Number of Visited Locations (N_LOC), similar to lower time scales, the data show some variances based on the choice of hexagonal bin resolution and the stay duration threshold. Table 4-6 illustrates that an increase in the bin resolution, while keeping the stay duration constant, results in an increase in the average number of visited locations. This could signify that individuals are more likely to be categorized as visiting distinct places when the spatial granularity is refined. Conversely, increasing the stay duration while keeping the bin size constant leads to a decline in N_LOC. This observation might indicate that individuals' movements appear less dispersed over larger time windows. Figure 4-53 and Figure 4-54 augment these findings by displaying the PMF and CDF curves, which show slight variations in the N_LOC peaks and upper limits, particularly for different stay duration thresholds.

Similarly, the Number of Unique Visited Locations (N_ULOC) also varies as the hexagonal bin resolution increases as shown in Table 4-7. Mirroring the results observed in lower time scales for both N_LOC and N_ULOC. Figure 4-55 and Figure 4-56 corroborate these observations. It's important to note that despite the increases, the number of unique locations (N_ULOC) consistently remains below 120, even at the highest resolutions. The PMF curves across different sampling thresholds remain consistent, signifying robustness in these metrics even when the observation scales change.

The Number of Significant Locations (N_SIG), presented in Table 4-8, remains surprisingly consistent across different hexagonal bin resolutions and stay duration thresholds. One possible explanation for this could be that regardless of how finely we parse the spatial or temporal data, individuals tend to frequent a core set of locations. Figure 4-57 further supports this conclusion, showing a striking similarity

in the PMF curves across all the different resolutions and stay duration thresholds examined.

The Radius of Gyration (R_GYR) showcases a level of stability across different hexagonal bin resolutions, as reflected in Table 4-9. The PMF curves in Figure 4-58 further support this, except for minor variations for R_GYR values below 30 km at the h7 resolution. This consistency suggests that the parameter is relatively insensitive to the choice of spatial granularity. The same observation is made in lower time scales.

Distance Traveled (D_TRAV), on the other hand, shows noticeable variability with changing hexagonal bin resolutions as shown in Table 4-10. This could imply that the total distance traversed by individuals is sensitive to how the geographic space is partitioned. Figure 4-59 supports this, indicating similar PMF curves but with differences that may be consequential, underscoring the need for careful consideration in selecting appropriate spatial resolutions for such analyses.

Finally, the Travel Time Percentage (T_TP) offers further nuanced insights. Table 4-11 shows that individuals spend more time traveling as the duration threshold increases. This is corroborated by Figure 4-60 and Figure 4-61, which although consistent in PMF shape, exhibit minor but noteworthy variations in the peaks of T_TP across different duration thresholds. The data suggest that people's travel time allocations are influenced both by the duration of stays at locations and the resolution of the spatial grid employed. In summary, the findings suggest that certain mobility parameters demonstrate a consistent pattern not only on a monthly scale but also on daily and weekly scales. This temporal consistency across different scales underscores the reliability and robustness of the parameters analyzed, further strengthening the validity of the study's conclusions.

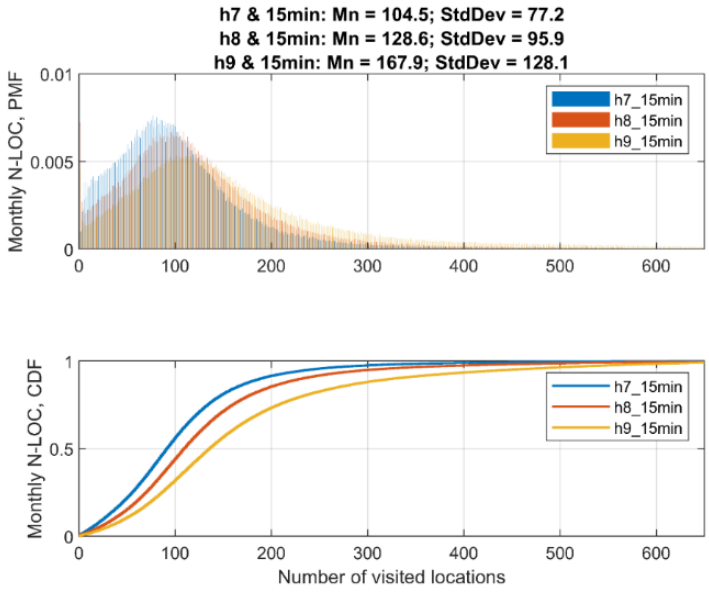

Figure 4-53: Estimated PMF and CDF for N_LOC monthly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 2020).

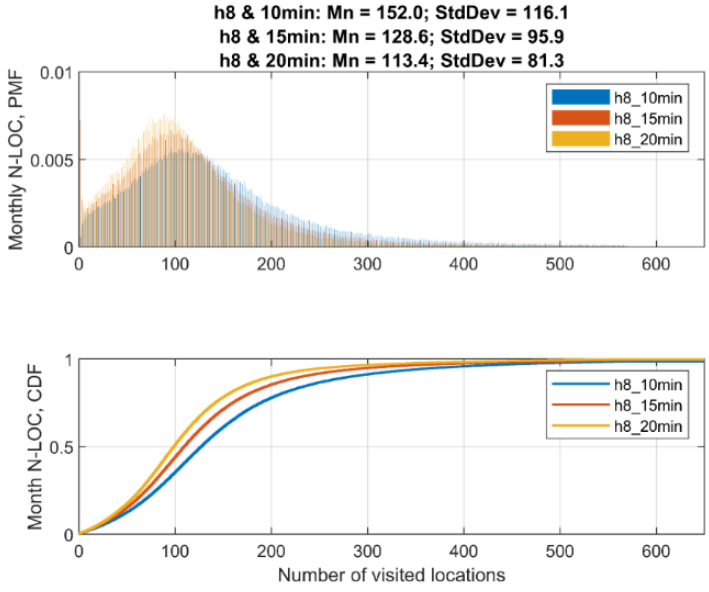

Figure 4-54: Estimated PMF and CDF for N_LOC monthly: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 2020).

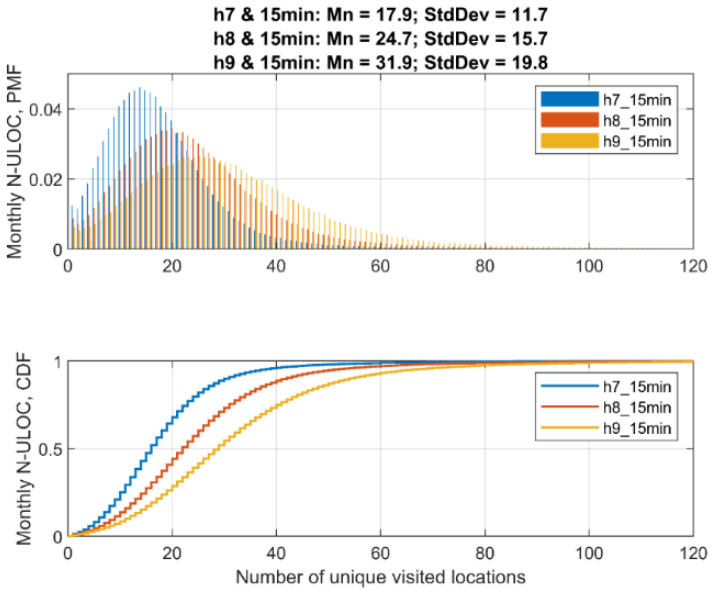

Figure 4-55: Estimated PMF and CDF for N_ULOC monthly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 2020).

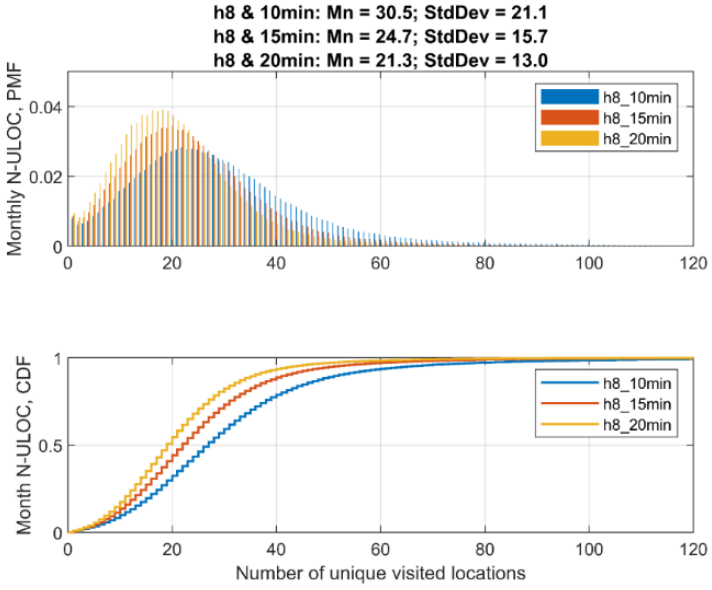

Figure 4-56: Estimated PMF and CDF for N_ULOC monthly: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (Oct. 2020).

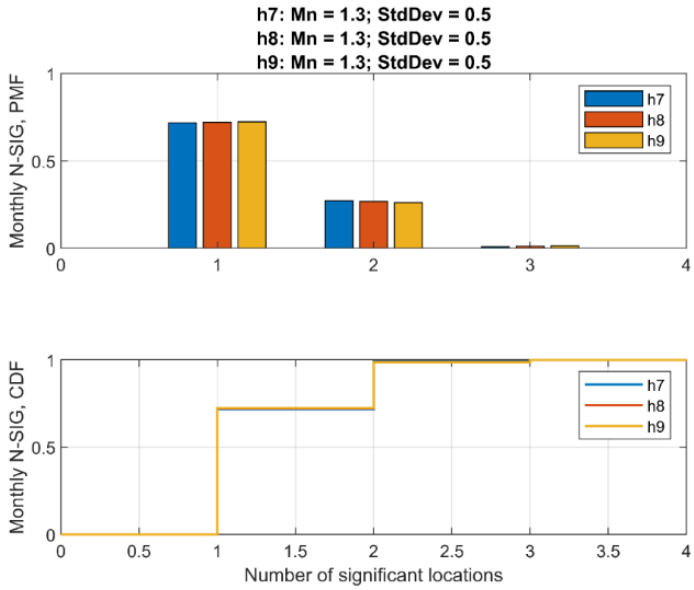

Figure 4-57: Estimated PMF and CDF for N_SIG monthly: resolution h7, h8, and h9 (October 2020).

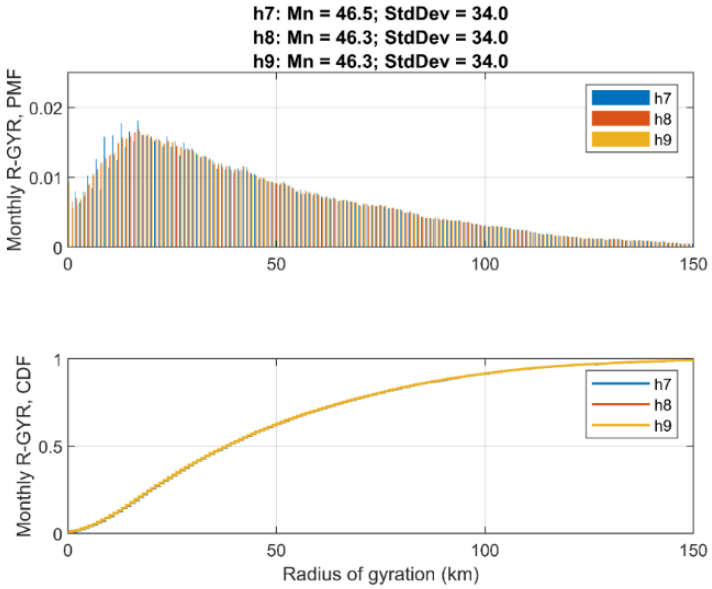

Figure 4-58: Estimated PMF and CDF for R_GYR monthly: resolution h7, h8, and h9 (October 2020).

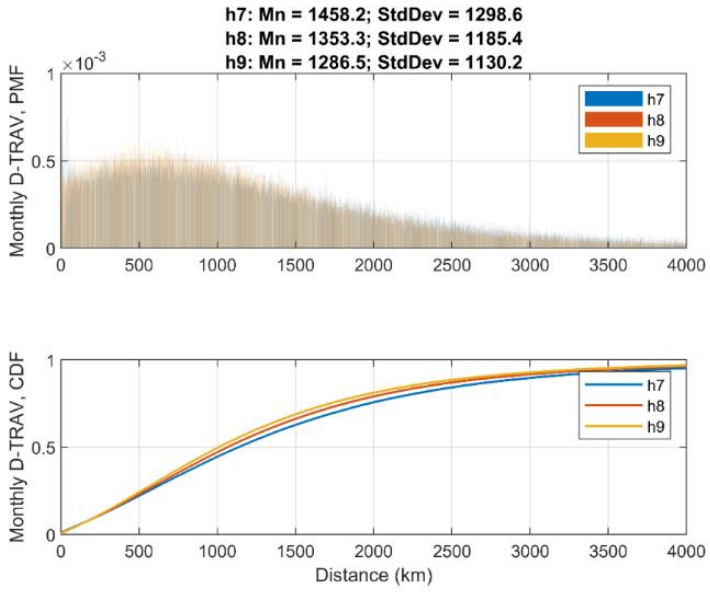

Figure 4-59: Estimated PMF and CDF for D_TRAV monthly: resolution h7, h8, and h9 (October 2020).

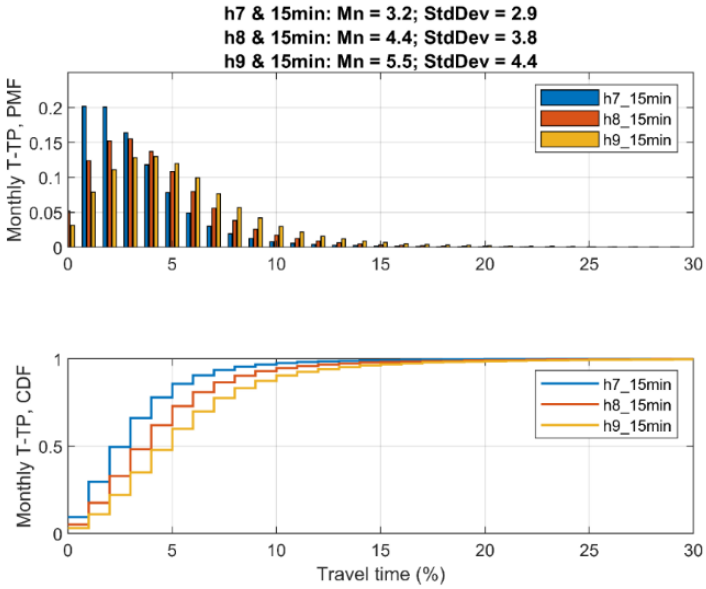

Figure 4-60: Estimated PMF and CDF for T_TP monthly: resolution h7, h8, and h9 for a stay duration threshold of 15 min (October 2020).

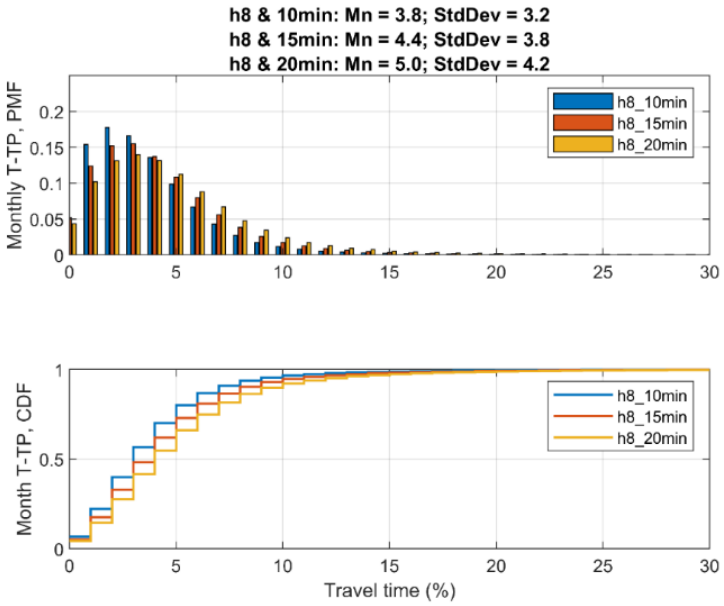

Figure 4-61: Estimated PMF and CDF for T_TP monthly: resolution h8 for a stay duration threshold of 10 min, 15 min, and 20 min (October 2020).

Comparison of Results with Related Work in the Literature

In this section, the results from the evaluation of the six mobility parameters are compared with relevant studies in the literature. This comparison aims to highlight both the similarities and differences between the current findings and previous research. Additionally, potential reasons for these observations will be discussed.

Number of Visited Locations (N_LOC)

In the research conducted by C. Song et al. [3], two user trajectories were analyzed to assess the number of visited locations. The first user visited approximately 22 locations per month within a 30 km region, while the second

frequented 76 locations per month spanning roughly a 90 km neighborhood. The "location" in their study is designated by a mobile phone tower, each covering an average reception area of 3 km2. This area represents the inherent uncertainty in pinpointing a user's exact location. Given the expansive coverage of a single location in their study, the most comparable results from our research correspond to the lowest uber h3 resolution, specifically h7, which equates to a radius of 1.12 km. Song et al. did not specify a stay duration threshold to determine if a location was visited. Therefore, a comparison will be made using a 15-minute stay duration from our dataset. For the Houston MSA, individuals visit an average of 121.7 locations per month at the h7 resolution with a 15-minute stay duration, and there's a standard deviation of 78.6. For the NY-NJ-PA MSA, the figure stands at 104.53 locations with a standard deviation of 77.15. Both figures are substantially greater than those presented in [3]. This discrepancy might stem from the difference in area coverages. Moreover, the geographic placement of the towers in [3] remains undisclosed, and the study may have been conducted outside the United States.

Ashbrook et al. [62] focused on a singular user, tracking their movements over four months, primarily within Atlanta, Georgia. Their data logger recorded the GPS receiver's output at one-second intervals, but only when the receiver was moving at a minimum speed of one mile per hour. The GPS's accuracy was gauged at 15 meters. They presented a graph detailing the number of places visited based on varying time thresholds. To align with our research parameters, we glean that at a 15-minute threshold, the user visited around 600 locations in [62]. Our results that can be most closely matched with this study are from the h9 resolution in our research, which is associated with a radius of 176 meters. For the Houston MSA, the average number of locations visited stands at 181.1 per month for h9 resolution with a 15-minute duration, and the standard deviation is 126.8. For the NY-NJ-PA MSA, the figure is 167.9 with a standard deviation of 128.1. The discrepancy between our findings and

[62] might arise from the h9 resolution's size, which is substantially larger than the 15 meters indicated in Ashbrook et al.'s research. In summation, our results align more closely with [62] than with [3].

Number of Unique Visited Locations (N_ULOC)

In [63], the datasets under review originate from three distinct continents. The Ivory Coast dataset is anchored on cellular locations, whereas those from India and Switzerland integrate Wi-Fi locations. There's a conspicuous variation in mobility patterns across these nations, with the Swiss populace visiting more places daily compared to their counterparts. The numbers from India and Switzerland exceed those obtained in this current research. One plausible explanation for this discrepancy could be the temporal threshold applied in this study. A salient observation from [63] is that individuals from developing countries tend to frequent fewer regular places compared to those from developed nations. Upon comparing the place-visiting patterns between India and Switzerland, several similarities surface, including the median duration of stay and the preference for Saturdays for less frequent visits. Nonetheless, distinctions in peak visit times were evident.

Shifting focus to [64], the study covers an 18-month duration starting in late 2009. It engaged 114 volunteer participants, each of whom utilized the designated smartphone as their sole mobile device. Predominantly, participants were between the ages of 20-40, and they spanned professionals as well as students from two educational institutions. On an average scale, each participant provided data for about 14 months. However, there were intervals, accounting for about 17% of the days, when the devices were switched off, leading to non-recording periods. This dataset has its roots in Switzerland.

In [64], a "place" is characterized as a circular area with a 100-meter radius that has been visited for a significant time. This conceptualization is in close proximity to the h9 resolution in the current study, which has a reach of 176 meters. The [64] researchers estimate that, on average, a participant ventured to 90 distinct locations over the research period. Their data suggests that a substantial majority visited between 50-150 places. Interestingly, 8% of the participants explored more than 150 locales, while 16% ventured to less than 50. It's pivotal to note that there was variability in the recording durations among participants, spanning between 4 to 18 months, which might have impacted the results. [64] recorded a weekly average of 7.5 unique locations visited, a figure that is slightly reduced compared to the findings of this research. Factors like the stay duration threshold and bin size employed in this study could account for such disparities.

Radius of Gyration (R_GYR)

In [15], W. Sun et al. examine an anonymous dataset sourced from a leading 4G cellular network provider in China. This dataset encompasses around 400,000 4G users, detailing their extensive HTTP records across three days in April 2015 from a city in northwest China. The study notes that a substantial 80% of users' daily radii of gyration are confined within a 10 km area. In contrast, a smaller segment (10%) cover distances beyond 30 km. Another research [65] points out that about 34% of all users have a radius of gyration less than 10 km, while only 14% have a radius exceeding 500 km. Directly juxtaposing these numbers with the findings of the current research is challenging due to the distinct three-day span considered in [15]. However, a brief glance at Table 4-3 for the Houston MSA and Table 4-9 for the New York-Newark-Jersey City MSA suggests that the R_GYR values for both MSAs significantly surpass those documented for the Chinese city in both [15] and [65].

In [66], the study delves into the daily ranges, specifically the radii of gyration, for two major US cities: Los Angeles and New York. Seasonal variations in these metrics are also evaluated. In New York, they identify average daily R_GYRs of 4 km during winter, which rise to 6 km in summer. Conversely, for Los Angeles, the daily R_GYRs average at 6 km in winter, peaking at 8 km in the summer season. Notably, these metrics are slightly below the values identified for both Houston and New York MSAs in this research.

Distance Traveled (D_TRAV)

The U.S. Department of Transportation [67] routinely compiles statistics reflecting driving volumes within the country. These statistics, available online, present average annual miles driven both nationally and for individual states. For the context of this study, the driving data from Texas and New York are of particular interest. In Texas, the annual driving distance, denoted as D_TRAV, averages around 16,000 miles. This breaks down to approximately 1,333 miles or roughly 2,145 kilometers monthly. Interestingly, these figures are slightly higher than those derived from the current research across all resolutions. One possible explanation for this disparity lies in the methodology this research employs to compute D_TRAV. Here, distance is calculated based on a direct linear trajectory, which tends to undervalue the true travel distances. In contrast, New York's data showcases an annual D_TRAV of about 10,167 miles, translating to monthly figures of 847 miles or 1,363 kilometers. These numbers from New York resonate closely with the findings of this research. The website also mentions that on average, American drive 14,263 miles (22,954 km) per year according to the Federal Highway Administration.

Travel Time Percentage (T_TP)

The studies from different sources provide insights into driving behaviors across the United States. According to [68], the average American dedicates nearly

18 days annually to driving, equating to a T_TP value of approximately 4.93%. Similarly, [69] indicates that Americans spend an estimated 19 days each year stuck in traffic, translating to an average T_TP of 5.21%. Further corroborating these findings, [70] states that the typical American allocates 18 days each year to driving, averaging eight hours and 22 minutes weekly. This survey, involving 2,000 American car owners, not only gauged the time commitment to driving but also delved into the emotional bond drivers share with their vehicles.

However, it is essential to note a few caveats when interpreting these findings. First, the data presented in these studies generally reflect national averages and lack geographical specificity. Secondly, these studies predominantly target individuals for whom driving isn't a professional obligation, overlooking a significant demographic that drives as a core component of their occupation. In juxtaposition with these studies, this dissertation research reveals that the T_TP value for the New York MSA hovers around 5%, aligning closely with the previously mentioned figures.

Number of Significant locations (N_SIG)

In [71], a systematic approach is developed to identify significant locations, termed as Points of Interest (PoI). The relevance of these locations is derived from the ratio of the number of days a place is visited to the total days within the evaluation span. Subsequent to determining their significance, these PoIs are categorized into three distinct classifications: Most Visited Points (MVPs), Occasionally Visited Points (OVPs), and Exceptionally Visited Points (EVPs). A distinctive aspect of their study lies in linking the count and nature of these pivotal locations to an individual's characteristics. Their findings, in essence, can be seen as a continuation or deepening of the research discussed in this dissertation. It's noteworthy that the figures associated with MVPs and OVPs from their analysis closely resonate with the

statistical data delineated in this current study, indicating potential consistencies in human movement patterns across diverse datasets.

Implications of the Findings

The findings from this study hold substantial implications in several sectors, including urban planning, transportation, and social sciences.

In the realm of social sciences, the data collected can offer valuable insights to scholars interested in urban human behavior. The observed variations in mobility patterns between weekdays and weekends could highlight aspects of lifestyle, work-life balance, and social activities within the population.

From the perspective of health and environmental implications, the comprehension of mobility patterns has the potential to help in predicting and managing disease spread. By understanding how and where people move, it is feasible for health authorities to anticipate disease transmission hotspots more effectively, thereby devising more targeted intervention strategies. Additionally, the evaluation and reduction of environmental effects such as pollution and greenhouse gas emissions could be accomplished more efficiently by understanding these mobility trends.

In the area of urban planning and policy formulation, insights into mobility patterns, specifically the count of visited locations and unique visited locations, offer a crucial understanding of urban space dynamics. Urban planners may utilize these insights to efficiently allocate public resources, cityscape design, and enhance urban infrastructures. Law makers could employ this knowledge to construct regulations in line with the observed movement trends, consequently improving the quality of urban life.

In the context of transportation management, an in-depth examination of mobility parameters such as the radius of gyration, distance traveled, and the percentage of travel time could substantially influence transportation strategies. These insights could be leveraged by transportation authorities to optimize public transportation routes, schedules, and capacity management, thereby improving the overall commuting experience, and reducing travel times.

As for machine learning and predictive modeling, these insights could assist in enhancing predictive models related to human mobility, contributing to areas such as traffic forecasting, emergency preparedness, and resource allocation. The incorporation of the spatial and temporal characteristics of human mobility into machine learning models could result in more precise and robust predictions.

In summary, the findings from this study provide a detailed interpretation of human mobility patterns. However, these insights should be interpreted within a broader context, considering factors like social and economic factors, cultural dynamics, and other anomalies. The ongoing research and analysis on mobility data should ultimately develop a more comprehensive understanding of human mobility, enabling more informed decision-making across different sectors.

Chapter 5
Comparative Study in 15 Top MSAs

This chapter focuses on the 15 largest Metropolitan Statistical Areas (MSAs) in the United States, selected based on population estimates for the year 2020. The objective of this study is to evaluate and statistically compare the results from four primary spatial parameters: Number of Visited Locations (N_LOC), Number of Unique Visited Locations (N_ULOC), Radius of Gyration (R_GYR), and Distance Traveled (D_TRAV) Furthermore, additional derived mobility parameters are also considered in this study. This research builds upon the methodology and results presented in Chapter 4. The study follows the data preprocessing steps outlined in Chapter 3 – Data Preprocessing to select unique mobile devices. This approach ensures the inclusion of only those devices that appear in the dataset every day of the month and record a minimum of 24 location entries per day, or at least one every hour. Table 3-1 presents the population size, the count of unique mobile devices data from which are included in the dataset, and the number of binned locations incorporated in this study. As previously mentioned, the data are geographically binned using Uber's H3 Hexagonal Hierarchical Spatial Index. The number of locations listed in Table 3-1 is determined based on the h8 resolution level, corresponding to an approximate hexagonal bin radius of 531 meters.

Mobility Parameters

After following the data preprocessing steps outlined in Chapter 3 – Data Preprocessing, four mobility parameters are evaluated for each of the studied Metropolitan Statistical Areas (MSAs): Number of visited locations (N_LOC), Number of unique visited locations (N_ULOC), Radius of gyration (R_GYR), and Distance traveled (D_TRAV). The individuals are considered to have visited a

location if they spend more than 15 minutes within the area associated with that location. Upon data preprocessing, the resultant dataset contains mean and standard deviations of the four studied mobility parameters for each day in the month of October 2020 and for each of the 15 identified metropolitan areas in the United States. In this chapter, the following additional parameters are derived for all studied areas:

Coefficients of Variation (CV): The coefficients of variation are calculated by dividing the standard deviation for a particular day with the mean for that specific day. This calculation presents the daily variation in D_TRAV, N_LOC, N_ULOC, and R_GYR as proportions of the mean, rather than in raw units of measurement or counts. Coefficients of variance are calculated for each of the four studied parameters.

Travel Path Shape: This parameter is calculated by dividing the mean distance traveled for a particular day with the radius of gyration for that specific day. This calculation helps determine if travel paths tend to more closely resemble a straight line than a circular shape. Higher values of the shape of travel parameter indicate that the path is more circular around the mobility center of mass.

Average Distance Between Locations: This parameter is calculated by dividing the mean distance traveled by the number of locations visited. This calculation shows the average distance covered between each visited location.

Statistical Analysis

The methodology presented in Chapter 3 and Chapter 4 is expanded upon in the present study, with the implementation of several additional analytical approaches, as outlined below:

- *Evaluation of Deviation from Theoretical Normal Distribution*: Measures of vertical and horizontal deviation, namely skewness and kurtosis, are analyzed to assess the extent of deviation from the ideal normal distribution. Deviation is determined in accordance with the standard rule of thumb: deviations of approximately ± 1 are deemed negligible and are therefore, disregarded [72].
- *Comparison of Daily Value Means*: The means of daily values concerning the mobility parameters under investigation across diverse MSAs are compared through the utilization of Repeated Measures Analysis of Variance (Repeated Measures ANOVA). The measure of effect size is represented by η^2.
- *Exploratory Factor Analysis among parameters within MSA*: This technique is employed to identify clusters of highly correlated mobility parameters. The identification of parameters with a strong correlation is conducted, indicative of their potential to provide identical information, hence leading to analogous results when incorporated in analyses. The exploratory factor analysis is carried out utilizing the principal axis factoring method, accompanied by a varimax orthogonal rotation of the final factor solution. The determination of the number of factors to extract is aided by Horn's parallel analysis [73].
- *Exploratory Factor Analysis among MSAs*: Exploratory factor analysis is used to create soft categories of MSAs according to daily mobility patterns. The factor analysis is applied to the same mobility parameters, measured in different MSAs on the same days. The data for this operation consist of

selected mobility parameters across all MSAs, with each MSA's data serving as one variable, and the different recording days being represented as entities/cases. Three distinct analyses are conducted based on mean daily distances travelled, mean daily number of locations visited, and the mean daily radius of gyration. Anderson-Rubin factor scores [74] of extracted factors are generated, subsequently used for further comparison to represent the generalized tendencies of the MSA group associated with that particular factor.

- *Factor Score Mean Values (daily comparisons)*: Mean values on different factor scores, which represent distinct groups of MSAs, are compared using one-way ANOVA. Eta squared (η^2) is utilized as a measure of effect size.
- *Pairwise Comparison of Means*: Following variance analysis, pairwise comparisons of means are conducted, applying the Bonferroni correction for probability inflation [75]. All statistical significances of pairwise comparisons include the Bonferroni correction. The statistical significance threshold employed in this study is 0.05. Cohen's D values are interpreted in accordance with the recommendations given by Cohen [76].
- *Examination of Autocorrelations*: Correlations between a parameter and its lagged values (lag 7 and lag 1) are calculated, serving to authenticate the observations about weekly cycle trends. Autocorrelations are derived by pairing parameter values with their own values at future time intervals. Here, the values are matched with the values of the same parameter on the corresponding weekday of the ensuing week, thereby validating the weekly cycle. This is contrasted with lag 1 autocorrelation, i.e., correlations of parameter values on each day, paired with their values on the following day, aiming to determine whether the weekly cycle is stronger than mere day-to-day comparisons.

Results and Discussion

A) Normality Test and ANOVA Results

Table 5-1 illustrates the mean, standard deviation, skewness, and kurtosis values for each mobility parameter evaluated during the period from October 1, 2020, to October 30, 2020. For comparative purposes, mean values across Metropolitan Statistical Areas (MSAs) are analyzed using repeated measures ANOVA, with eta squared (η^2) serving as a measure of effect size. To enhance readability, MSAs are referred to by the primary part of their name. For instance, the New York-Newark-Jersey City, NY-NJ MSA is referred to as New York MSA.

A review of Table 5-1 suggests that the distribution of the majority of assessed parameters closely aligns with a theoretical normal distribution, as most skewness and kurtosis values lie within the ± 1 range [42]. Noteworthy deviations are observed in the shapes of daily travel paths, which demonstrate considerable negative asymmetry, and daily average distances between locations, which are either normally distributed for 6 MSAs or positively skewed for 9 MSAs.

Despite some parameters exhibiting distributions deviating beyond the ± 1 range, the choice was made to retain parametric statistics without implementing normalization transformations to the data. This decision was informed by the identical outcomes that would be derived from both parametric and nonparametric procedures in this context, especially considering the significant effect sizes (η^2) noted in Table 5-1.

The data in Table 5-1 shows notable contrasts in the mean daily distance traveled across MSAs such as Dallas, Houston, and Atlanta where average daily distances span 61 km to 63 km, and in MSAs like San Francisco, Los Angeles, and Seattle, which record shorter distances of 37 km, 43 km, and 43 km respectively. Standard deviations of daily travel distance are consistently higher than mean values,

indicating a considerable range in individual travel patterns. This suggests the presence of many individuals with minimal or no travel (thus reducing the mean) as well as a smaller subset of individuals with extensive travel (thus significantly increasing the standard deviation).

Evaluation of the number of locations visited reveals that the MSAs with the most extensive and least extensive daily travel distances also recorded the highest and lowest average numbers of locations visited, respectively. A similar trend is noted when assessing the number of unique locations visited. However, the three MSAs with the smallest average daily number of unique locations visited are San Francisco, Seattle, and Washington, with Riverside and Los Angeles recording slightly higher values.

Assessment of the average radius of gyration shows that Dallas, Houston, and Atlanta MSAs have the highest values, while San Francisco, Los Angeles, and Seattle MSAs have the lowest. In terms of the average distance between locations visited, data indicates that visited locations are furthest apart in Houston, Dallas, and Atlanta MSAs, and closest together in San Francisco, New York, and Los Angeles MSAs.

The mean values of daily travel paths suggest that these paths are generally more linear than circular, with relatively low standard deviations. Among different areas, Riverside MSA exhibits paths most closely aligned with a linear shape, while Houston, Detroit, and Dallas MSAs show the most deviation from this shape, suggesting a more circular traffic infrastructure in these MSAs.

Comparison of mean values of the studied parameters between the fifteen (15) MSAs using repeated measures ANOVA reveals that differences between MSAs are statistically significant in all cases and of extreme size (high η^2 values). Within each MSA, differences among parameter values on different days of October 2020 tend to

be smaller than those among different MSAs. The smallest difference among MSAs, though still notably large, is seen when examining mean coefficients of variation of the daily N_LOC ($\eta^2 = 17$), while the largest difference is found when comparing mean values of daily average distance between locations ($\eta^2 = 0.96$).

This suggests that the number of locations people visit within MSAs does not vary as significantly as the average distances individuals travel. The MSAs are more similar when considering the frequency of location visits, but they diverge more substantially in terms of the average distances covered by individuals.

B) Exploratory Factor Analysis

A preliminary pairwise comparison among the 15 MSAs is conducted, resulting in a correlation matrix that encompasses daily values of the studied mobility parameters for each MSA. Upon inspection of this matrix, it becomes evident that the differences in mean values for these parameters are not universally statistically significant across all MSAs. Specifically, the mean value of a given parameter in one MSA often does not differ significantly from that in several other MSAs, although it may stand out in the remaining ones. This suggests that grouping MSAs with similar mean values on the same days is feasible. Further examination of the daily parameter values within each MSA reveals that most of these parameters exhibit a strong degree of correlation. These correlations can be either positive or negative, and occasionally, no significant correlation is identified. The complexity and size of these correlation matrices make them unsuitable for either interpretation or presentation within a research paper. As a result, an exploratory factor analysis is first performed on parameters that relate to the same MSA.

Table 5-1: Descriptive Statistics and ANOVA Results for Different MSAs

Mean Daily D_TRAV (km)				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	61.88	5.7	-0.44	0.36
Boston	45.51	4.74	1.05	0.76
Chicago	48.53	4.69	0.23	0.02
Dallas	62.8	6.15	-0.02	-1.08
Detroit	52.62	5.1	-0.33	-0.18
Houston	62.83	5.19	0.01	-0.76
Los Angeles	42.68	3.41	0.75	-0.85
Miami	47.12	4.14	-0.69	0.39
New York	44.91	4.68	0.81	0.69
Philadelphia	45.2	4.89	0.78	0.01
Phoenix	56.46	5.2	-0.54	-0.04
Riverside	49.57	3.81	-0.19	-0.66
San Francisco	37.44	3.08	0.96	-0.22
Seattle	43.32	3.74	-0.53	0.05
Washington	46.52	5.09	0.66	0.14
F	**323.48**			
$\boldsymbol{\eta^2}$	**0.92**			
Sig.	**<0.01**			

Mean SD of Daily D_TRAV (km)				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	66.57	3.31	-0.91	1.28
Boston	53.62	2.95	1.14	0.86
Chicago	59.15	2.57	-0.46	0.72
Dallas	68.67	3.62	-0.38	-0.89
Detroit	57.64	2.87	-0.69	0.36
Houston	68.82	2.83	-0.23	-0.76
Los Angeles	52.57	2.06	1.04	0.08
Miami	55.58	2.66	-1.05	0.86
New York	57.16	3.54	0.98	0.31
Philadelphia	51.67	2.46	0.54	-0.44
Phoenix	65.34	3.15	-0.78	0.15
Riverside	61.73	2.24	-0.18	-0.45
San Francisco	49.44	1.7	0.34	-0.97
Seattle	54.69	2.57	-1.2	1.14
Washington	58.14	2.87	0.28	0.07
F	**448.81**			
$\boldsymbol{\eta^2}$	**0.94**			
Sig.	**<0.01**			

Mean Daily N_LOC				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	5.61	0.35	-0.58	0.77
Boston	5.1	0.31	0.12	-0.84
Chicago	5.17	0.32	-0.44	0.41

Mean Daily SD of N_LOC				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	4.24	0.1	2.88	10.89
Boston	3.91	0.06	-0.33	-0.26
Chicago	4	0.06	-1.07	1.34

Mean Daily N_LOC (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Dallas	5.69	0.36	-0.56	-0.02
Detroit	5.26	0.34	-0.78	0.66
Houston	5.66	0.32	-0.76	0.79
Los Angeles	4.87	0.23	0.38	-0.92
Miami	5.25	0.31	-0.77	0.55
New York	5.17	0.34	-0.49	0.19
Philadelphia	5.12	0.34	-0.3	-0.57
Phoenix	5.2	0.31	-0.58	0.5
Riverside	4.82	0.23	-0.13	-0.23
San Francisco	4.56	0.21	0.64	-0.88
Seattle	4.75	0.25	-0.96	0.98
Washington	4.79	0.32	-0.42	0.13
F	**128.96**			
η^2	**0.82**			
Sig.	**<0.01**			

Mean Daily SD of N_LOC (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Dallas	4.19	0.07	-0.64	-0.09
Detroit	3.96	0.07	-0.64	-0.03
Houston	4.17	0.06	-0.91	0.86
Los Angeles	3.81	0.04	-0.65	-0.13
Miami	3.98	0.07	-1.26	1.17
New York	4.03	0.06	-0.68	-0.15
Philadelphia	3.93	0.06	-0.94	0.14
Phoenix	3.89	0.08	-0.96	0.8
Riverside	3.85	0.06	-0.77	0.74
San Francisco	3.77	0.05	-0.8	0.84
Seattle	3.95	0.07	-1.25	1.66
Washington	3.94	0.07	-1.61	1.92
F	**331.08**			
η^2	**0.92**			
Sig.	**<0.01**			

Mean Daily N_ULOC				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	3.15	0.2	0.02	0.49
Boston	2.84	0.18	0.53	-0.43
Chicago	2.9	0.19	0.14	-0.37
Dallas	3.16	0.21	0.06	-0.83
Detroit	2.95	0.2	-0.12	-0.42
Houston	3.17	0.19	0.12	-0.46
Los Angeles	2.79	0.15	0.74	-0.96

Mean Daily SD of N_ULOC				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	1.91	0.11	-0.22	0.68
Boston	1.7	0.08	0.14	-0.49
Chicago	1.78	0.09	-0.49	0.4
Dallas	1.91	0.11	-0.26	-0.48
Detroit	1.77	0.1	-0.49	0.08
Houston	1.91	0.1	-0.27	-0.22
Los Angeles	1.73	0.07	0.26	-0.69

Mean Daily N_ULOC (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Miami	3.03	0.18	-0.21	0.01
New York	2.92	0.2	0.02	-0.07
Philadelphia	2.86	0.2	0.27	-0.67
Phoenix	2.97	0.18	0.04	-0.21
Riverside	2.75	0.14	0.41	-0.63
San Francisco	2.62	0.14	0.89	-0.66
Seattle	2.7	0.15	-0.14	-0.31
Washington	2.74	0.2	0.31	-0.38
F	**119.34**			
η^2	**0.8**			
Sig.	**<0.01**			

Mean Daily SD of N_ULOC (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Miami	1.89	0.1	-1.02	1.24
New York	1.8	0.08	-0.22	-0.42
Philadelphia	1.71	0.09	-0.31	-0.57
Phoenix	1.85	0.1	-0.62	0.53
Riverside	1.7	0.07	-0.39	0.31
San Francisco	1.66	0.07	0.53	-0.69
Seattle	1.74	0.09	-0.78	0.94
Washington	1.74	0.1	-0.65	0.54
F	**172.9**			
η^2	**0.86**			
Sig.	**<0.01**			

Mean Daily R_GYR (km)				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	12.56	1.08	0.94	0.16
Boston	9.46	1.33	1.16	0.45
Chicago	10.06	1.12	1.17	0.55
Dallas	12.72	1.34	0.9	-0.13
Detroit	10.58	0.98	1.11	0.15
Houston	12.61	1.2	1.02	-0.14
Los Angeles	8.94	0.9	1.01	-0.09
Miami	9.52	0.76	1.12	0.39
New York	9.4	1.39	0.93	-0.46
Philadelphia	9.27	1.17	1.19	0.68
Phoenix	11.42	0.93	1.02	-0.45

Mean Daily SD of R_GYR (km)				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	14.09	1.21	0.8	-0.5
Boston	11.74	1.72	1.03	-0.35
Chicago	12.74	1.4	0.95	-0.39
Dallas	14.57	1.44	0.64	-1.06
Detroit	11.64	0.93	0.88	-0.56
Houston	14.52	1.47	0.64	-1.31
Los Angeles	11.33	1.23	0.86	-1
Miami	12.1	1.29	0.94	-0.57
New York	13.2	2.14	0.82	-1
Philadelphia	10.79	1.29	0.89	-0.85
Phoenix	13.74	1.2	0.73	-0.86

Mean Daily R_GYR (km) (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Riverside	11.19	0.97	0.93	-0.33
San Francisco	7.97	0.84	1.02	0.06
Seattle	9.18	0.78	1.17	0.47
Washington	9.89	1.25	1.21	0.43
F	**295.91**			
η^2	**0.91**			
Sig.	**<0.01**			

Mean Daily SD of R_GYR (km) (Continued)				
MSA	Mean	SD	Skewness	Kurtosis
Riverside	16.97	2.07	0.63	-0.7
San Francisco	10.24	0.88	0.86	-0.94
Seattle	11.68	0.92	0.99	-0.14
Washington	12.53	1.35	0.93	-0.64
F	**323.42**			
η^2	**0.92**			
Sig.	**<0.01**			

Mean Daily D_TRAV CV				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	1.08	0.05	0.42	0.6
Boston	1.18	0.06	-0.4	-0.44
Chicago	1.22	0.07	-0.22	-0.26
Dallas	1.1	0.05	-0.12	-1.08
Detroit	1.1	0.06	0.41	-0.05
Houston	1.1	0.05	0.06	-0.68
Los Angeles	1.24	0.05	-0.5	-1.07
Miami	1.18	0.05	0.7	0.55
New York	1.28	0.06	0.38	0.76
Philadelphia	1.15	0.07	-0.4	-0.05
Phoenix	1.16	0.06	0.61	0.15
Riverside	1.25	0.05	0.39	-0.56
San Francisco	1.33	0.06	-0.92	-0.18
Seattle	1.27	0.06	0.15	-0.43
Washington	1.26	0.07	-0.42	-0.07

Mean Daily N_LOC CV				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	0.76	0.04	1.14	1.69
Boston	0.77	0.04	0.1	-0.67
Chicago	0.78	0.04	0.61	0.55
Dallas	0.74	0.04	0.81	0.12
Detroit	0.75	0.04	1.13	1.46
Houston	0.74	0.03	0.92	0.63
Los Angeles	0.78	0.03	-0.37	-0.93
Miami	0.76	0.03	0.72	0.35
New York	0.78	0.04	0.88	1.52
Philadelphia	0.77	0.04	0.36	-0.52
Phoenix	0.75	0.03	0.72	0.61
Riverside	0.8	0.03	0.01	-0.83
San Francisco	0.83	0.03	-0.54	-0.71
Seattle	0.83	0.03	0.73	0.31
Washington	0.82	0.04	0.24	0.07

Mean Daily D_TRAV CV (Continued)				
F	**199.17**			
η^2	**0.87**			
Sig.	**<0.01**			
Mean Daily N_ULOC CV				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	0.61	0.01	0.37	0.22
Boston	0.6	0.02	0.08	-0.54
Chicago	0.62	0.01	-0.62	-0.34
Dallas	0.61	0.01	0.13	0.07
Detroit	0.6	0.01	-0.53	-0.63
Houston	0.6	0.01	-0.1	-0.1
Los Angeles	0.62	0.01	-0.16	-0.99
Miami	0.62	0.01	0.05	-0.43
New York	0.62	0.02	0.24	0.57
Philadelphia	0.6	0.01	-0.47	0.27
Phoenix	0.62	0.01	0.3	-0.79
Riverside	0.62	0.01	-0.32	-0.98
San Francisco	0.64	0.01	-0.3	-0.59
Seattle	0.65	0.01	-0.46	0.03
Washington	0.64	0.02	-0.79	-0.27
F	**104.98**			
η^2	**0.78**			
Sig.	**<0.01**			

Mean Daily N_LOC CV (Continued)				
F	**69.69**			
η^2	**0.71**			
Sig.	**<0.01**			
Mean Daily R_GYR CV				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	1.12	0.06	1.65	2.51
Boston	1.24	0.05	1.11	0.25
Chicago	1.27	0.06	1.68	2.87
Dallas	1.15	0.05	1.46	1.24
Detroit	1.1	0.06	1.79	2.81
Houston	1.15	0.05	1.92	2.49
Los Angeles	1.27	0.05	1.38	1.53
Miami	1.27	0.08	1.36	1.13
New York	1.4	0.07	1.57	1.87
Philadelphia	1.17	0.06	1.27	0.7
Phoenix	1.2	0.07	1.6	1.84
Riverside	1.51	0.1	1.35	0.63
San Francisco	1.29	0.05	-0.22	-0.84
Seattle	1.27	0.06	1.79	2.54
Washington	1.27	0.07	1.4	1.93
F	**292.42**			
η^2	**0.91**			
Sig.	**<0.01**			

Mean Daily Travel Path Shape				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	4.93	0.34	-1.16	-0.23
Boston	4.84	0.32	-1.08	-0.16
Chicago	4.84	0.31	-1.31	0.47
Dallas	4.95	0.35	-1.23	0.1
Detroit	4.98	0.31	-1.43	0.77
Houston	5	0.37	-1.26	0.1
Los Angeles	4.79	0.26	-1.27	0.16
Miami	4.96	0.36	-1.16	-0.05
New York	4.82	0.35	-1.15	0.01
Philadelphia	4.89	0.32	-1.32	0.62
Phoenix	4.95	0.31	-1.24	0.22
Riverside	4.44	0.28	-1.13	0.04
San Francisco	4.71	0.24	-1.27	0.15
Seattle	4.73	0.28	-1.22	0
Washington	4.72	0.28	-1.25	0.4
F	**161.95**			
$\boldsymbol{\eta^2}$	**0.85**			
Sig.	**<0.01**			

Mean Daily Average Distance Between Locations (km)				
MSA	Mean	SD	Skewness	Kurtosis
Atlanta	11.02	0.49	-0.11	0.68
Boston	8.91	0.51	1.38	1.27
Chicago	9.37	0.39	1.09	0.61
Dallas	11.02	0.51	1.08	0.18
Detroit	9.98	0.39	0.69	0.08
Houston	11.09	0.43	1.19	0.36
Los Angeles	8.75	0.34	1.26	0.76
Miami	8.97	0.32	-0.28	0.09
New York	8.69	0.55	1.12	-0.09
Philadelphia	8.81	0.45	1.51	1.56
Phoenix	10.83	0.43	0	-0.73
Riverside	10.28	0.34	0.19	-0.25
San Francisco	8.2	0.33	1.1	0.72
Seattle	9.1	0.35	0.24	-0.22
Washington	9.69	0.48	1.52	1.43
F	**603.24**			
$\boldsymbol{\eta^2}$	**0.96**			
Sig.	**<0.01**			

This approach aims to identify clusters of parameters that deliver similar or almost identical information, thereby eliminating the need for separate analysis. Examples of the results from this exploratory factor analysis are illustrated in Tables 5-2 and Table 5-3, focusing on the mobility parameters values associated with the Atlanta MSA and the Boston MSA, respectively. The method used for this exploratory factor analysis is principal axis factoring. Factor loadings, acquired through varimax orthogonal rotation, are presented in the table. The decision on the number of factors to extract is guided by Horn's parallel analysis. Factor loadings that fall below 0.40 are considered low in this case and are not displayed for the sake of clarity.

Figure 5-1 displays scree plots that compare the eigenvalues of factors extracted from the parameters studied within the Atlanta MSA against factors extracted from a simulated dataset. This comparison falls within the scope of Horn's parallel analysis, which aims to determine the appropriate number of factors to extract.

Table 5-2: Atlanta MSA Mobility Parameters Factor Analysis Results

Parameter	Factor 1	Factor 2	Uniqueness
Mean D_TRAV	0.83	0.56	0.00
SD of D_TRAV	0.85	0.47	0.05
Mean N_LOC	0.96		0.03
SD N_LOC	0.52		0.72
Mean N_ULOC	0.86	0.48	0.04
SD N_ULOC	0.91		0.10
Mean R_GYR		0.98	0.01
SD of R_GYR		0.88	0.11
D_TRAV CV	-0.81	-0.59	0.00
N_LOC CV	-0.91		0.06
N_ULOC CV		-0.80	0.30
R_GYR CV	-0.95		0.09
Travel Path Shape	0.88	-0.44	0.04
Average Distance Between Locations	0.42	0.82	0.16

Table 5-3: Boston MSA Mobility Parameters Factor Analysis Results

Parameter	Factor 1	Factor 2	Uniqueness
Mean D_TRAV	0.71	0.70	0.00
SD of D_TRAV	0.55	0.81	0.05
Mean N_LOC	0.95		0.01
SD N_LOC	0.88		0.22
Mean N_ULOC	0.87	0.48	0.01
SD N_ULOC	0.90		0.16
Mean R_GYR		0.96	0.00
SD of R_GYR		1.00	0.00
D_TRAV CV	-0.86	-0.50	0.01
N_LOC CV	-0.91		0.06
N_ULOC CV	-0.44	-0.73	0.27
R_GYR CV	-0.89		0.14
Travel Path Shape	0.49	-0.87	0.00
Average Distance Between Locations		0.92	0.08

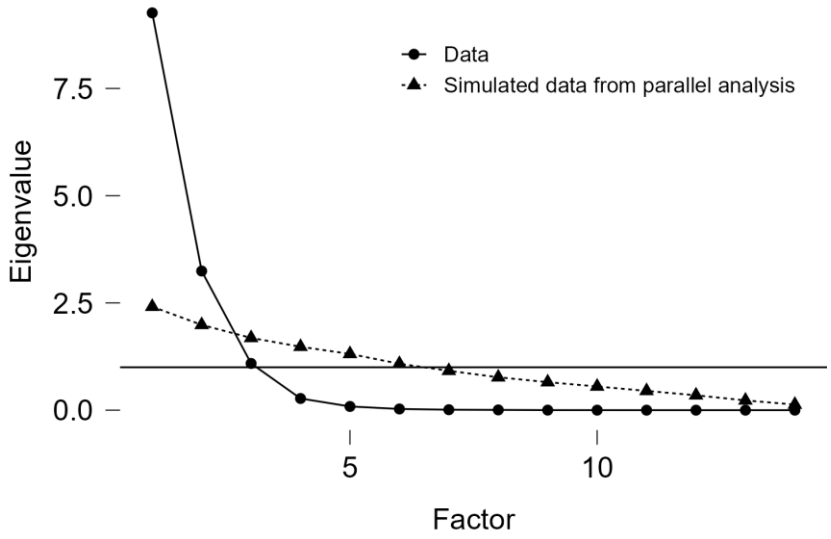

Figure 5-1: Atlanta MSA Scree Plot (Horn's Parallel Analysis).

The results from these analyses indicate that, typically, all parameters can be categorized into two distinct groups. One factor consistently includes the mean and standard deviation of the radius of gyration as well as the average distance between locations. On the other hand, the parameters that represent the number of locations and the number of unique locations constitute the second factor. Interestingly, the mean distance travelled often displays a relatively high correlation with both of these factors.

It is worth noting that both the standard deviation of the number of locations visited and the coefficient of variation of the number of unique locations visited are parameters that show the highest level of uniqueness in both the Atlanta and Boston MSAs.

Factor analysis is then conducted to identify clusters of MSAs. Based on the previous analysis results, the primary grouping is established based on the average daily distance traveled. Further groupings are analyzed according to the average number of daily locations visited and the radius of gyration.

The results of the exploratory factor analysis conducted on the mean distance traveled across different MSAs are illustrated in Table 5-4. Additionally, Figure 5-2 presents a scree plot that shows the eigenvalues resulting from this analysis, alongside the eigenvalues of factors from a simulated matrix. It should be noted that factor loadings below 0.58 are not included in Table 5-4.

Table 5-4: Factor Loadings and Uniqueness of Daily Mean Distance Traveled.

Mean Daily Distance Travelled			
MSA	Factor 1	Factor 2	Uniqueness
Atlanta	0.87		0.20
Boston		0.92	0.12
Chicago	0.62	0.67	0.18
Dallas	0.82		0.08
Detroit	0.82		0.10
Houston	0.87		0.05
Los Angeles	0.67	0.71	0.06
Miami	0.89		0.19
New York		0.98	0.02
Philadelphia		0.84	0.07
Phoenix	0.95		0.01
Riverside	0.90		0.01
San Francisco		0.72	0.16
Seattle	0.89		0.10
Washington		0.79	0.11

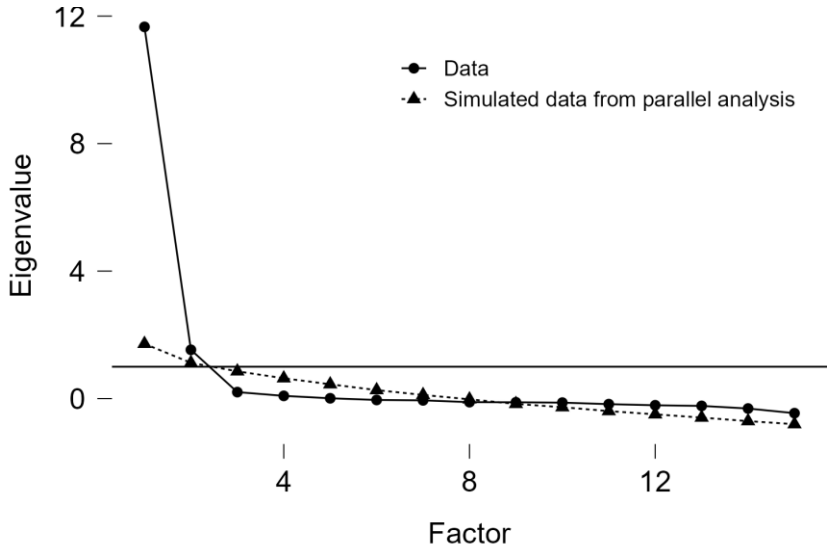

Figure 5-2: Scree plot of D_TRAV extracted factors with eigenvalues in different MSAs.

From Table 5-4 results, two factors are extracted. These factors collectively represent 90% of the variance in daily mean distance traveled, specifically, 52.3% and 38.1% respectively. The examination of the results indicates that the first factor (Factor 1) is characterized by high loadings in the Atlanta, Dallas, Detroit, Houston, Miami, Phoenix, Riverside, and Seattle MSAs. Similarly, the second factor (Factor 2) exhibits high loadings in Boston, Los Angeles, New York, Philadelphia, San Francisco, and Washington. Interestingly, Chicago and Los Angeles stand out by displaying high loadings on both factors.

MSAs with high loadings on Factor 1 are predominantly found to be inland, with the exceptions of Miami and Seattle. In contrast, all MSAs with high loadings on Factor 2, barring Chicago, are identified as coastal cities. These findings might illustrate the distinctive patterns of daily distance traveled between inland and coastal cities, allowing for a few outliers.

When these results are compared with the average daily mean distance traveled, MSAs with the highest mean values of distance traveled, such as Dallas, Houston, and Atlanta, are all found to display high loadings on Factor 1. Seattle, on the other hand, which has one of the lowest distances traveled values, also displays a high loading on Factor 1. Conversely, San Francisco and Los Angeles, the other two low mean daily distance traveled MSAs, display high loadings on Factor 2, although Los Angeles also exhibits a significant loading on Factor 1. These observations might suggest that the patterns of daily distance traveled changes responsible for this grouping are not highly associated with the average size of daily distance traveled, considering that the factor analytic procedure is conducted on the correlation matrix and correlations are not sensitive to differences in means and variances.

For the purpose of corroborating these findings, the same procedure is repeated on the mean daily number of locations visited and the mean daily radius of gyration across different MSAs, in light of the preceding results.

Table 5-5 shows the Factor loadings and uniqueness of the daily mean number of locations visited for various MSAs. Factor loadings below 0.58 are not displayed.

In Figure 5-3, the scree plot illustrates the extracted factors from the mean daily number of locations visited, along with the corresponding eigenvalues derived from a simulated matrix created under the framework of Horn's parallel analysis.

Upon examining the factor analysis results for the daily mean number of locations visited, it is found that the outcomes again consist of 2 factors. These 2 factors together explain 85.2% of the variance in the number of locations visited, specifically 51.7% and 33.4%. However, it is notable that the second factor demonstrates high loadings for the daily number of locations visited data from New York, Philadelphia, Boston, and Washington, while the remaining MSAs exhibit the highest loading on the first factor. Chicago exhibits equally high loadings on both factors, similar to the results observed for mean distances traveled.

The total variance explained by these factors is somewhat lower compared to the analysis based on mean distances traveled. Consequently, the average uniqueness of source variables in this analysis is higher than in the one conducted on mean distance travel, approximately 0.15 here compared to 0.10 in the analysis on mean distance traveled. The uniqueness of mean daily numbers of locations visited for San Francisco, Miami, and Chicago stands out with 34%, 26%, and 25% of unique variance, respectively.

Table 5-5: Factor Loadings and Uniqueness of Daily Mean N_LOC

Mean Daily N_LOC			
MSA	Factor 1	Factor 2	Uniqueness
Atlanta	0.88		0.20
Boston		0.87	0.21
Chicago	0.60	0.63	0.25
Dallas	0.82		0.14
Detroit	0.78		0.15
Houston	0.90		0.07
Los Angeles	0.77		0.14
Miami	0.83		0.26
New York		0.92	0.07
Philadelphia		0.90	0.04
Phoenix	0.94		0.06
Riverside	0.89		0.03
San Francisco	0.65		0.34
Seattle	0.83		0.13
Washington		0.79	0.14

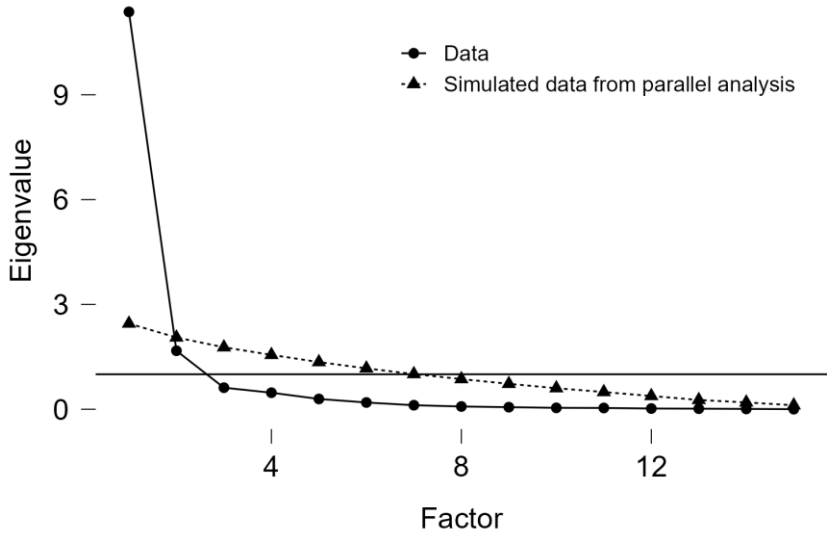

Figure 5-3: Scree plot of N_LOC extracted factors with eigenvalues in different MSAs.

A closer examination of the MSAs highly loading on the second factor reveals that they are all situated on the U.S. east coast and adjacent to one another along the north-south axis. On the other hand, Factor 1 MSAs are located either inland or on the west or southern part of the U.S., with Chicago being an exception, as it is located inland and exhibits moderate loadings on both factors.

The results of the factor analysis for the mean daily radius of gyration of the MSAs under study are presented in Table 5-6 and Figure 5-4. Factor loadings below 0.58 are not displayed. In contrast to the previous analysis, Horn's parallel analysis indicates that all mean daily radius of gyration variables for the studied MSAs form a single factor, which accounts for 86% of the variance in daily values across all MSAs. Highly significant loadings on this factor are observed for all MSAs.

Boston, Atlanta, and Miami stand out due to their relatively high levels of uniqueness in their daily mean radius of gyration values. However, since only one factor was extracted in this analysis, the daily radius of gyration of different cities cannot be utilized to classify them into distinct groups. Instead, they all belong to the same group, as they are characterized by the same underlying factor.

C) Day of the Week Factor Scores and Comparisons

Table 5-7 presents the descriptive statistical analysis of Anderson-Rubin factor scores extracted from mean daily distance travelled, mean daily number of locations visited, and mean daily radius of gyration on different days of the week. The table includes means, standard deviations, standard errors of means, and 95% confidence intervals of means for the mentioned factors. Upon examining the results, it is evident that Factors 1 and 2 scores exhibit differences mainly concerning the day of the week when they reach their peak values. Specifically, Factor 1 scores demonstrate the highest levels on Fridays, whereas Factor 2 scores are most prominent on Saturdays.

Table 5-6: Factor Loadings and Uniqueness of Daily Mean R_GYR

Mean Daily R_GYR		
MSA	Factor 1	Uniqueness
Atlanta	0.85	0.28
Boston	0.82	0.33
Chicago	0.92	0.15
Dallas	0.98	0.04
Detroit	0.94	0.11
Houston	0.99	0.03
Los Angeles	0.99	0.02
Miami	0.88	0.22
New York	0.86	0.26
Philadelphia	0.95	0.09
Phoenix	0.92	0.15
Riverside	0.97	0.06
San Francisco	0.97	0.07
Seattle	0.91	0.17
Washington	0.96	0.09

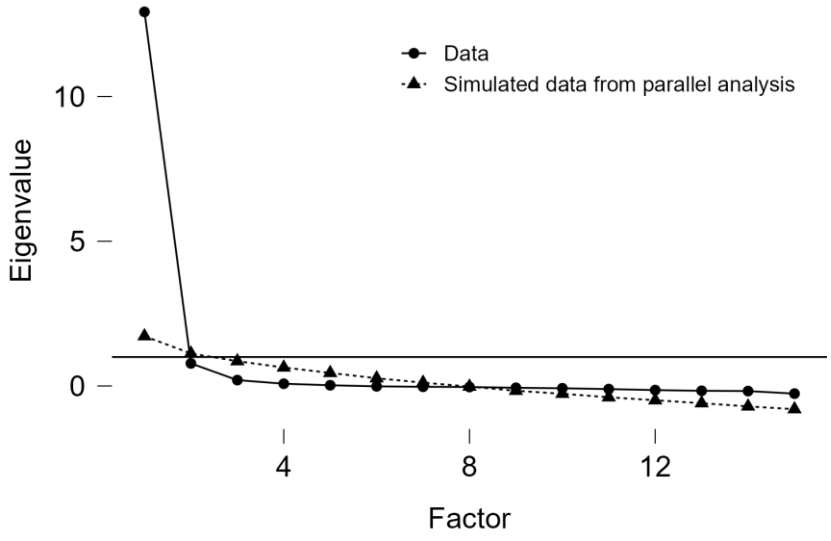

Figure 5-4: Scree plot of R_GYR extracted factors with eigenvalues in different MSAs.

Table 5-7: Descriptive Statistics of Factor Scores for Daily Travel and Visitation Patterns

Mean Daily D_TRAV – Factor 1 Scores Stats					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Monday	-0.32	0.23	0.11	-0.69	0.04
Tuesday	-0.12	0.20	0.10	-0.43	0.20
Wednesday	0.14	0.09	0.04	0.00	0.29
Thursday	0.30	0.33	0.15	-0.11	0.72
Friday	1.27	0.49	0.22	0.66	1.87
Saturday	0.41	0.31	0.16	-0.09	0.91
Sunday	-2.07	0.21	0.11	-2.41	-1.74
Mean Daily D_TRAV – Factor 2 Scores Stats					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Monday	-1.04	0.52	0.26	-1.87	-0.21
Tuesday	-0.63	0.26	0.13	-1.05	-0.22
Wednesday	-0.43	0.22	0.11	-0.77	-0.08
Thursday	-0.34	0.61	0.27	-1.09	0.42
Friday	0.15	0.56	0.25	-0.55	0.84
Saturday	2.04	0.31	0.16	1.55	2.54
Sunday	0.29	0.41	0.21	-0.36	0.95
Mean Daily N_LOC – Factor 1 Scores Stats					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Monday	-0.28	0.37	0.19	-0.88	0.31
Tuesday	-0.13	0.27	0.14	-0.56	0.31
Wednesday	0.11	0.26	0.13	-0.30	0.52
Thursday	0.29	0.70	0.31	-0.58	1.16
Friday	1.37	0.71	0.32	0.48	2.25
Saturday	0.07	0.23	0.12	-0.30	0.44
Sunday	-1.85	0.23	0.12	-2.21	-1.48
Mean Daily N_LOC – Factor 2 Scores Stats					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Monday	-0.86	1.02	0.51	-2.49	0.76
Tuesday	-0.21	0.42	0.21	-0.87	0.46
Wednesday	-0.10	0.40	0.20	-0.73	0.53

Mean Daily N_LOC – Factor 2 Scores Stats (Continued)					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Thursday	-0.13	1.17	0.53	-1.59	1.33
Friday	0.34	1.00	0.45	-0.90	1.58
Saturday	1.50	0.25	0.13	1.10	1.91
Sunday	-0.60	0.54	0.27	-1.46	0.26
Mean Daily R_GYR – Factor 1 Scores Stats					
Day	Mean	Standard Deviation	SE of mean	95% CI of mean, lower and upper bound	
Monday	-0.96	0.12	0.06	-1.15	-0.77
Tuesday	-0.84	0.07	0.04	-0.96	-0.73
Wednesday	-0.65	0.08	0.04	-0.78	-0.53
Thursday	-0.47	0.05	0.03	-0.56	-0.39
Friday	0.66	0.29	0.13	0.30	1.02
Saturday	2.05	0.10	0.05	1.89	2.21
Sunday	0.16	0.15	0.07	-0.08	0.40

As the week progresses, scores for both factors generally experience an increasing trend, reaching their highest values on a specific day and then subsequently declining.

Factor 1 scores demonstrate a distinct pattern of being notably lower, particularly on Sundays. Conversely, Factor 2 scores show their lowest levels on Mondays. Consequently, these findings suggest a classification of the studied MSAs into two groups based on individuals travel and location visitation patterns. The first group, Factor 1 MSAs, corresponds to regions where individuals tend to travel extensively and visit numerous locations on Fridays. In contrast, the second group, Factor 2 MSAs, comprises areas where people engage in substantial travel and visit multiple locations on Saturdays. Furthermore, it is evident that Factor 1 MSAs experience significantly reduced travel on Sundays, whereas Factor 2 MSAs still demonstrate comparatively higher travel distances on Sundays, albeit with fewer location visits.

Regarding mean daily locations visited, Factor 1 MSAs are characterized by a higher frequency of visits on Fridays, followed by minimal or limited travel on Sundays. On the other hand, Factor 2 MSAs exhibit an elevated number of location visits on Saturdays, but subsequently show a decrease in visits on Sundays. A possible explanation for these findings might be differences in shopping behaviors among the two groups. Specifically, it is possible that individuals in Factor 1 MSAs conduct their shopping activities predominantly on Fridays, leading to multiple location visits on that day. Conversely, Factor 2 MSAs might be exhibiting shopping patterns primarily focused on Saturdays, resulting in a significant number of locations visited on Saturdays. Additionally, mean radius of gyration displays its highest values on Saturdays and the lowest values on Mondays. The values progressively increase from Monday to Saturday, before decreasing again on Sundays.

It is crucial to emphasize that while these findings provide valuable insights into the weekly cycle trends, comprehensive research is necessary for a thorough understanding of the observed differences. Importantly, the dataset under examination does not provide sufficient detail to ascertain the precise nature of these travel patterns. Traveling on Fridays and Saturdays could conceivably be influenced by other factors, such as social interactions or recreational pursuits. In the absence of additional research into the motivations behind these mobility patterns, attributing them solely to shopping or any other specific activity remains speculative.

Table 5-8 presents the ANOVA results, comparing mean values for different days of the week based on factor scores from Table 5-8. The analysis indicates significant differences in factor score means across the days for all factors. High η^2 values indicate that these differences between days are very pronounced. However, when considering the results for Factor 2 that are derived from the average daily number of visited locations the effect size is significant but not as pronounced.

Table 5-8: ANOVA Results for Mean Values Across Days of the Week on Extracted Factors.

Factor score	F	Statistical Sig.	η^2
Factor 1 mean daily D_TRAV	48.26	<0.01	0.93
Factor 2 mean daily D_TRAV	19.92	<0.01	0.84
Factor 1 mean daily N_LOC	17.78	<0.01	0.82
Factor 2 mean daily N_LOC	3.74	0.01	0.49
Factor 1 mean daily R_GYR	200.16	<0.01	0.98

D) Pairwise Comparison of Factor Scores Means

Pairwise comparisons of MSAs factor scores means for different days of the week are conducted, and the Bonferroni correction is used to account for the inflation of probabilities due to multiple comparisons. Sizes of differences between means are expressed using Cohen's d-s. Since data is collected from a single month, statistical significance for differences between means is achieved mostly for the most pronounced discrepancies. This data limitation, along with the application of the Bonferroni correction, results in only the most substantial differences meeting the usual thresholds for statistical significance.

In Table 5-9, pairwise comparisons of mean factor scores across different days of the week for Factor 1 MSAs, extracted from average daily distances traveled, are presented. A Bonferroni correction is applied, considering a family of 7. Differences with significant statistical prominence are highlighted in bold. The most significant differences, often reaching statistical significance, are observed between Friday and Sunday as compared to other days. This suggests that the mobility parameters on Fridays and Sundays are notably different from those observed on the remaining days. Similarly, in Table 5-10, pairwise comparisons for Factor 2 MSAs, based on average daily distances traveled, are displayed. The most significant disparities are

observed between Saturday and the other days in the week. This indicates that the average daily distance covered on Saturdays is substantially greater than on the remaining days of the week for Factor 2 MSAs.

Tables 5-11 and 5-12 show pairwise comparisons for Factors 1 and 2 MSAs, respectively, derived from the average daily number of locations visited. Table 5-11 reveals significant differences between Friday and Sunday compared to other days, with the greatest distinction observed between these two days. In Table 5-12, the most substantial difference was observed between Monday and Saturday.

While many differences were noteworthy according to Cohen's d-s values [46], only the variations between Saturdays, Sundays, and Mondays met the commonly accepted statistical significance threshold of 0.05. This suggests that in Factor 1 MSAs, individuals, on average, frequent the greatest number of locations on Fridays and the fewest on Sundays. Conversely, in Factor 2 MSAs, the highest number of locations are visited on Saturdays, and the lowest on Mondays.

Table 5-13 presents pairwise comparisons for Factor 1 MSAs based on the average daily radius of gyration. Upon applying the Bonferroni correction, it becomes evident that the differences between the means of nearly all days of the week are statistically significant and accompanied by pronounced effect sizes. The notable exceptions, though still significant in terms of effect size, were between consecutive days in the first half of the week and between Monday and Wednesday. This finding indicates a relatively consistent daily radius of gyration across different days of the week. Further analysis of the data reveals that for Factor 1 MSAs, the highest radius of gyration occurs on Fridays and the lowest on Sundays. In contrast, for Factor 2 MSAs, the peak radius of gyration is observed on Saturdays while the lowest is on Mondays.

Table 5-9: Mean D_TRAV Factor 1 Scores for Days of the Week With Bonferroni Correction

Factor 1 - Mean Daily D_TRAV						
Days	Weekdays	Mean Difference	SE of difference between means	t	Cohen's d	Statistical Sig. (Bonferroni correction)
Monday	Tuesday	-0.21	0.22	-0.95	-0.67	1.00
	Wednesday	-0.46	0.22	-2.15	-1.52	0.88
	Thursday	-0.62	0.20	-3.04	-2.04	0.12
	Friday	**-1.59**	**0.20**	**-7.77**	**-5.21**	**<0.01**
	Saturday	**-0.73**	**0.22**	**-3.41**	**-2.41**	**0.05**
	Sunday	**1.75**	**0.22**	**8.13**	**5.75**	**<0.01**
Tuesday	Wednesday	-0.26	0.22	-1.20	-0.85	1.00
	Thursday	-0.42	0.20	-2.04	-1.37	1.00
	Friday	**-1.38**	**0.20**	**-6.77**	**-4.54**	**<0.01**
	Saturday	-0.53	0.22	-2.45	-1.74	0.47
	Sunday	**1.96**	**0.22**	**9.08**	**6.42**	**<0.01**
Wednesday	Thursday	-0.16	0.20	-0.77	-0.52	1.00
	Friday	**-1.12**	**0.20**	**-5.50**	**-3.69**	**<0.01**
	Saturday	-0.27	0.22	-1.25	-0.89	1.00
	Sunday	**2.21**	**0.22**	**10.28**	**7.27**	**<0.01**
Thursday	**Friday**	**-0.97**	**0.19**	**-5.01**	**-3.17**	**<0.01**
	Saturday	-0.11	0.20	-0.55	-0.37	1.00
	Sunday	**2.37**	**0.20**	**11.61**	**7.79**	**<0.01**
Friday	Saturday	0.85	0.20	4.18	2.80	0.01
	Sunday	**3.34**	**0.20**	**16.34**	**10.96**	**<0.01**
Saturday	**Sunday**	**2.48**	**0.22**	**11.53**	**8.16**	**<0.01**

Table 5-10: Mean D_TRAV Factor 2 Scores for Days of the Week With Bonferroni Correction

Factor 2 - Mean Daily D_TRAV						
Days	Weekdays	Mean Difference	SE of difference between means	t	Cohen's d	Statistical Sig. (Bonferroni correction)
Monday	Tuesday	-0.40	0.32	-1.27	-0.89	1.00
	Wednesday	-0.61	0.32	-1.92	-1.36	1.00
	Thursday	-0.70	0.30	-2.32	-1.56	0.62
	Friday	**-1.18**	**0.30**	**-3.91**	**-2.62**	**0.02**
	Saturday	**-3.08**	**0.32**	**-9.65**	**-6.83**	**<0.01**
	Sunday	**-1.33**	**0.32**	**-4.18**	**-2.95**	**0.01**
Tuesday	Wednesday	-0.21	0.32	-0.66	-0.46	1.00
	Thursday	-0.30	0.30	-0.99	-0.66	1.00
	Friday	-0.78	0.30	-2.58	-1.73	0.35
	Saturday	**-2.68**	**0.32**	**-8.39**	**-5.93**	**<0.01**
	Sunday	-0.93	0.32	-2.91	-2.06	0.16
Wednesday	Thursday	-0.09	0.30	-0.30	-0.20	1.00
	Friday	-0.57	0.30	-1.89	-1.27	1.00
	Saturday	**-2.47**	**0.32**	**-7.73**	**-5.47**	**<0.01**
	Sunday	-0.72	0.32	-2.26	-1.60	0.71
Thursday	Friday	-0.48	0.29	-1.69	-1.07	1.00
	Saturday	-2.38	0.30	-7.86	-5.27	<0.01
	Sunday	-0.63	0.30	-2.08	-1.40	1.00
Friday	**Saturday**	**-1.90**	**0.30**	**-6.27**	**-4.20**	**<0.01**
	Sunday	-0.15	0.30	-0.49	-0.33	1.00
Saturday	**Sunday**	**1.75**	**0.32**	**5.48**	**3.87**	**<0.01**

Table 5-11: Mean N_LOC Factor 1 Scores for Days of the Week With Bonferroni Correction

Factor 1 - Mean Daily N_LOC						
Days	Weekdays	Mean Difference	SE of difference between means	t	Cohen's d	Statistical Sig. (Bonferroni correction)
Monday	Tuesday	-0.16	0.33	-0.47	-0.33	1.00
	Wednesday	-0.39	0.33	-1.18	-0.83	1.00
	Thursday	-0.58	0.32	-1.82	-1.22	1.00
	Friday	**-1.65**	**0.32**	**-5.20**	**-3.49**	**<0.01**
	Saturday	-0.35	0.33	-1.06	-0.75	1.00
	Sunday	**1.56**	**0.33**	**4.68**	**3.31**	**<0.01**
Tuesday	Wednesday	-0.24	0.33	-0.71	-0.50	1.00
	Thursday	-0.42	0.32	-1.32	-0.89	1.00
	Friday	**-1.49**	**0.32**	**-4.70**	**-3.15**	**<0.01**
	Saturday	-0.20	0.33	-0.59	-0.42	1.00
	Sunday	**1.72**	**0.33**	**5.15**	**3.64**	**<0.01**
Wednesday	Thursday	-0.18	0.32	-0.58	-0.39	1.00
	Friday	**-1.26**	**0.32**	**-3.96**	**-2.66**	**0.01**
	Saturday	0.04	0.33	0.12	0.08	1.00
	Sunday	**1.96**	**0.33**	**5.85**	**4.14**	**<0.01**
Thursday	**Friday**	**-1.07**	**0.30**	**-3.58**	**-2.27**	**0.03**
	Saturday	0.22	0.32	0.70	0.47	1.00
	Sunday	**2.14**	**0.32**	**6.75**	**4.53**	**<0.01**
Friday	**Saturday**	**1.30**	**0.32**	**4.08**	**2.74**	**0.01**
	Sunday	**3.21**	**0.32**	**10.13**	**6.79**	**<0.01**
Saturday	**Sunday**	**1.92**	**0.33**	**5.73**	**4.05**	**<0.01**

Table 5-12: Mean N_LOC Factor 2 Scores for Days of the Week With Bonferroni Correction

Factor 2 - Mean Daily N_LOC						
Days	Weekdays	Mean Difference	SE of difference between means	t	Cohen's d	Statistical Sig. (Bonferroni correction)
Monday	Tuesday	-0.66	0.57	-1.16	-0.82	1.00
	Wednesday	-0.77	0.57	-1.36	-0.96	1.00
	Thursday	-0.74	0.54	-1.37	-0.92	1.00
	Friday	-1.20	0.54	-2.24	-1.51	0.73
	Saturday	**-2.37**	**0.57**	**-4.19**	**-2.96**	**0.01**
	Sunday	-0.26	0.57	-0.46	-0.33	1.00
Tuesday	Wednesday	-0.11	0.57	-0.20	-0.14	1.00
	Thursday	-0.08	0.54	-0.15	-0.10	1.00
	Friday	-0.55	0.54	-1.02	-0.68	1.00
	Saturday	-1.71	0.57	-3.03	-2.14	0.13
	Sunday	0.40	0.57	0.70	0.50	1.00
Wednesday	Thursday	0.03	0.54	0.06	0.04	1.00
	Friday	-0.44	0.54	-0.81	-0.55	1.00
	Saturday	-1.60	0.57	-2.83	-2.00	0.20
	Sunday	0.51	0.57	0.90	0.64	1.00
Thursday	Friday	-0.47	0.51	-0.92	-0.58	1.00
	Saturday	-1.63	0.54	-3.04	-2.04	0.12
	Sunday	0.48	0.54	0.89	0.60	1.00
Friday	Saturday	-1.16	0.54	-2.17	-1.46	0.85
	Sunday	0.94	0.54	1.76	1.18	1.00
Saturday	**Sunday**	**2.11**	**0.57**	**3.73**	**2.64**	**0.02**

Table 5-13: Mean R_GYR Factor 1 Scores for Days of the Week With Bonferroni Correction

Factor 1 - Mean Daily R_GYR						
Days	Weekdays	Mean Difference	SE of difference between means	t	Cohen's d	Statistical Sig. (Bonferroni correction)
Monday	Tuesday	-0.12	0.11	-1.11	-0.78	1.00
	Wednesday	-0.31	0.11	-2.83	-2.00	0.20
	Thursday	**-0.49**	**0.11**	**-4.52**	**-3.19**	**<0.01**
	Friday	**-1.62**	**0.10**	**-15.76**	**-10.57**	**<0.01**
	Saturday	**-3.01**	**0.11**	**-27.79**	**-19.65**	**<0.01**
	Sunday	**-1.12**	**0.11**	**-10.33**	**-7.31**	**<0.01**
Tuesday	Wednesday	-0.19	0.11	-1.73	-1.22	1.00
	Thursday	**-0.37**	**0.11**	**-3.41**	**-2.41**	**0.05**
	Friday	**-1.50**	**0.10**	**-14.60**	**-9.79**	**<0.01**
	Saturday	**-2.89**	**0.11**	**-26.68**	**-18.87**	**<0.01**
	Sunday	**-1.00**	**0.11**	**-9.23**	**-6.52**	**<0.01**
Wednesday	Thursday	-0.18	0.11	-1.68	-1.19	1.00
	Friday	**-1.31**	**0.10**	**-12.77**	**-8.57**	**<0.01**
	Saturday	**-2.71**	**0.11**	**-24.96**	**-17.65**	**<0.01**
	Sunday	**-0.81**	**0.11**	**-7.50**	**-5.30**	**<0.01**
Thursday	**Friday**	**-1.13**	**0.10**	**-11.00**	**-7.38**	**<0.01**
	Saturday	**-2.52**	**0.11**	**-23.27**	**-16.46**	**<0.01**
	Sunday	**-0.63**	**0.11**	**-5.81**	**-4.11**	**<0.01**
Friday	**Saturday**	**-1.39**	**0.10**	**-13.53**	**-9.08**	**<0.01**
	Sunday	**0.50**	**0.10**	**4.87**	**3.27**	**<0.01**
Saturday	**Sunday**	**1.89**	**0.11**	**17.46**	**12.34**	**<0.01**

E) Examination of Autocorrelations

To validate observations of weekly patterns, autocorrelations between daily values of factor scores from the preceding analysis are determined. Based on earlier findings pointing to a 7-day trend [20], autocorrelations with a 7-day lag are calculated. For a clearer contrast, 1-day lag autocorrelations are also calculated, suggesting that the 7-day pattern provides a more insightful interpretation of daily changes than mere adjacent day comparisons. These findings are detailed in Table 5-14.

Table 5-14: Autocorrelations of Factor Scores With Lag7 and Lag1

Factor score	Lag7 autocorrelation	Lag1 autocorrelation
Factor 1 MSAs Mean Daily D_TRAV	0.94	0.16
Factor 2 MSAs Mean Daily D_TRAV	0.81	0.33
Factor 1 MSAs Mean Daily N_LOC	0.84	0.28
Factor 2 MSAs Mean Daily N_LOC	0.38	0.29
Factor 1 MSAs Mean Daily R_GYR	0.99	0.38

Upon examination of Table 5-14, it is observed that 7-day lag autocorrelations are consistently much higher compared to the 1-day lag. This implies that weekly cycles offer a better understanding of daily changes than mere day-to-day comparisons. However, an outlier is noted in Factor 2, derived from the average daily number of locations visited (N_LOC). For this factor, the 7-day lag autocorrelation is noticeably subdued compared to other factors and is only marginally above the 1-day lag, indicating a weaker 7-day trend for this specific factor.

Implications of the Findings

The observed statistical variances in daily mobility parameters across MSAs indicate the presence of spatial variability. It is inferred that movement patterns might be swayed by geographical or socio-economic considerations. In particular, different regions, whether inland or coastal, are identified to possess unique mobility features.

In the realm of data reduction and efficiency, the pronounced correlations among the mobility parameters within an MSA have been noted. Such correlations permit researchers and policymakers to narrow their focus to a curated set of parameters, ensuring that vital information isn't compromised. This approach not only simplifies the analytic process but also optimizes the use of resources.

When observing the geographic mobility traits, it's clear that the classifications arising from the factor analysis provide critical insights. Distinctions, especially between inland and coastal MSAs or the east coast versus other areas, suggest a significant impact of geographical or regional socio-cultural and economic determinants on mobility patterns. These insights bear great potential for endeavors in urban planning and infrastructural development.

Furthermore, the mobility data uncovers a consistent weekly pattern. This regularity in movement is of consequence to multiple sectors:

In transportation planning, distinguishing between peak and non-peak days can guide refinements in public transport scheduling, roadwork, and traffic control.

For the business sector, such rhythmic patterns allow businesses, from retailers to service providers, to refine their services, operational hours, or promotional efforts in sync with peak mobility days.

On the policy front, governmental bodies can harness this data, formulating policies that resonate with the established mobility trends, ensuring both efficient implementation and heightened public adherence.

The singular dominant factor discerned in the daily radius of gyration presents intriguing possibilities for additional research. Questions arise regarding the predominant influence of this factor and the underlying mechanisms that grant it such significance.

Should the focus of future research encompass health or safety, these mobility patterns become paramount. Grasping these patterns can bolster efforts in disease containment, planning for emergencies, or even framing health awareness initiatives.

In essence, the data shows interesting patterns of mobility across varied MSAs. These insights, while foundational for urban planners, policymakers, businesses, and researchers, truly come to life when integrated with the overarching aims and context of the study.

Chapter 6
Conclusion and Future Work

Conclusion

Through a comprehensive analysis of four mobility parameters in the Houston MSA region and six in the NY-NJ-PA MSA region, valuable insights into various aspects of human mobility patterns have been garnered. The results and discussions stemming from these analyses are detailed in Chapter 4.

The study reveals that an increase in the stay duration threshold results in a decrease in the number of visited and unique visited locations (N_LOC and N_ULOC respectively). These patterns hold true across all time scales: daily, weekly, and monthly. Conversely, an increase in the hexagonal bin resolution levels leads to an increase in these location parameters, indicating a significant relationship between these parameters and the temporal and spatial scales of analysis.

Despite the variations in other parameters, the radius of gyration (R_GYR) remains consistent across different hexagonal bin resolutions and is not influenced by changes in the stay duration threshold. This consistency is maintained across all temporal scales. It is worth mentioning that many individuals are found to have a zero R_GYR, signifying a stationary location throughout the observed period.

In relation to the distance traveled (D_TRAV), the study exposes minor variations that are dependent on the hexagonal bin resolution. While D_TRAV tends to fluctuate on a daily basis, it shows greater stability over weekly and monthly time scales. Confirming that the weekly time scale is the fundamental timescale of human mobility. To achieve a higher level of precision in distance calculations, the study recommends employing higher binning resolutions.

The average travel time percentage (T_TP) shows a marked dependence on both the stay duration threshold and hexagonal bin resolution. As these factors increase, so does the T_TP across all temporal scales, though fluctuations are minor on a weekly basis and only slightly more pronounced on a monthly scale.

Interestingly, the study also identifies a contrast between weekday and weekend mobility patterns. It is observed that weekdays typically see a higher number of visited locations and unique visited locations, while weekends have greater average distances traveled and travel times, along with a larger radius of gyration.

For the number of significant locations (N_SIG) visited, it is found that their numbers increase at higher resolution levels, but they remain unaffected by the stay duration threshold. This pattern remains consistent across daily and weekly timeframes, with monthly evaluations showing only a minor standard deviation.

The probability mass function (PMF) curves for all parameters exhibit a remarkable degree of consistency across different resolutions and thresholds. This uniformity in distribution patterns is indicative of the predictability of human mobility. The study thus suggests that with a broader range of studies across various periods and geographic regions, precise models can be developed. Furthermore, the study validates the use of cellular network data for analyzing human mobility patterns.

The choice of appropriate sampling thresholds, however, is not straightforward. It is dependent on the study's objective, the dataset size, and computational capacity. For parameters such as R_GYR, D_TRAV, and N_SIG, stay duration threshold holds no relevance, and the primary consideration should be the geographical binning resolution levels.

In summary, this comprehensive analysis presented in Chapter 4 explains the complex relationship between spatial resolution (hexagonal bin resolution), temporal

granularity (stay duration threshold), and human mobility patterns. The study also reveals that while different sampling threshold combinations may yield different means and standard deviations, the distribution, as seen in their PMF curves, remains strikingly similar. This study shows meaningful results that can lay the foundation for further studies to characterize human mobility.

In Chapter 5, a comprehensive analysis of spatial mobility parameters in 15 most populous MSAs is presented. Upon comparing the mean values of daily mobility parameters across different MSAs, statistically significant differences emerge. Notably, substantial effect sizes are evident across all parameters. When conducting a factor analysis of daily values of various mobility parameters within the same MSA, significant correlations among them are observed. As a result, individual analyses for each parameter are deemed inefficient and not warranted. Hence, emphasis in subsequent analyses is placed on three key mobility parameters: Mean Daily D_TRAV, which shows a strong correlation with others, and Mean Daily N_LOC and R_GYR, both of which demonstrate substantial loadings on factors derived from travel pattern parameters.

It is imperative to note that mobility parameters that load highly on the same factor inherently correlate with one another. This correlation implies not only that one parameter can be predicted from the others with notable accuracy, but also that executing separate analyses on each parameter would yield analogous results.

Further factor analysis on the daily mean distance travelled across different MSAs uncovers two primary factors that meet the parallel analysis criteria. Soft categorizations of MSAs are then created based on these factors: the first factor predominantly comprises inland MSAs, and the second mainly includes coastal MSAs. A similar two-factor structure emerges when analyzing the mean daily number of locations visited, with one factor aligning with east coast MSAs and the

other encompassing MSAs situated inland or in the southern or western regions of the US. A noteworthy observation in the factor analysis of the daily radius of gyration points to the existence of just one dominant factor, accounting for most of the variance in daily differences due to its high correlations among MSAs.

Investigating further, a pronounced weekly cycle trend emerges. MSAs classified under Factor 1 display their longest mean travel distances on Fridays and the shortest on Sundays. In contrast, those under Factor 2 travel the furthest on Saturdays and less so on Sundays, with Monday being their least mobile day. A similar pattern is identified in the number of daily locations visited by both groups: Factor 1 MSAs reach their peak activity on Fridays and show their lowest levels on Sundays, while Factor 2 MSAs reach their peak on Saturdays and minimal activity on Mondays.

Limitations and Future Work

Chapter 4 and Chapter 5 present a broad yet targeted study into human mobility, with a focus that is currently constrained to a limited set of parameters. It uses data from October 2020, a period characterized by COVID-19 lockdown measures, a factor that significantly impacted human mobility patterns. Thus, while the insights gained from this data provide valuable information for understanding human behavior under such exceptional circumstances, they may not fully capture the regular dynamics of human mobility in the region. This presents a key limitation of the study, as well as a potential direction for future research.

The scope of the analysis presented in Chapter 4 was narrowed to four mobility parameters in Houston MSA region and six key mobility parameters in the NY-NJ-PA MSA region, chosen due to their prevalence in existing literature. The extension of this research could include the exploration of additional parameters, potentially encompassing variables such as the average speed of movement, the number of trips taken, and the length of these trips. By adhering to the research methodology

developed in this study, it is hoped that a diverse and comprehensive set of mobility parameters can be established over time. This would allow for more robust comparisons between different geographical regions, timeframes, and population groups.

In terms of geographical context, this study concentrates on a large metropolitan area within the United States. The findings are likely applicable to similar urban environments. Nonetheless, the United States features a variety of geographical morphologies, including urban, suburban, and rural areas. Each of these regions likely exhibits unique mobility patterns. Hence, a comparative analysis investigating mobility differences across these diverse morphologies could yield valuable insights.

Broadening the scope further, the role of geographical location on human mobility is another area that warrants further investigation. It is plausible that mobility patterns observed in international cities may diverge from those documented within the United States. Therefore, undertaking comparative studies of mobility trends across various global regions could further enrich our understanding of human mobility.

The temporal scope of the data used in this research spans only the month of October 2020, potentially limiting the study's capacity to understand seasonal variations in human mobility. It is reasonable to suggest that mobility trends may be affected by factors such as school holidays, special events, weather conditions, or even natural disasters. Using the current research framework, future studies could make a meaningful and quantitative comparison of seasonal mobility trends.

Lastly, the process of collecting mobility data is often impractical, making the development of theoretical models an area of potential exploration. It is anticipated that accurate models of human mobility could result from these data collection and analysis efforts. Such models could then be utilized to predict human mobility under

a range of hypothetical scenarios. This capability could have a significant impact, particularly in the areas of policy decision-making, infrastructure development, and traffic management planning.

Chapter 5 embarks on an initial exploration into mobility patterns within the 15 most populous Metropolitan Statistical Areas (MSAs) in the U.S., recognizing several potential areas for intensifying insights and methodologies through deeper investigation.

One pivotal avenue for future work involves examining the causes of differences and similarities in mobility patterns across various MSAs. Investigating why mobility parameters in certain MSAs are similar to each other, while differing from MSAs in other groups, presents a critical area for exploration. This research suggests turning initially to census data and other structural characteristics of these MSAs and communities as valuable starting points for this investigative process. Such an inquiry could yield findings with significant theoretical and scientific contributions.

The merit of integrating data from diverse sources such as social media platforms, transportation networks, and advanced traffic monitoring systems is recognized. Through such integration, a multifaceted perspective on urban mobility is expected to emerge, potentially revealing intricate details that might be overlooked when relying solely on mobile location data.

Significance is also attributed to external determinants, such as climatic fluctuations or global health events, on mobility patterns. Understanding these factors is considered crucial for informing future urban planning and preparedness strategies.

Additional avenues include a more in-depth investigation into behavioral analytics. By aligning the acquired mobility data with sociological research, deeper insights into the underlying motivations behind movement patterns are expected to be gained.

References

[1] C. Kang et al., "Analyzing and geo-visualizing individual human mobility patterns using mobile call records," in 2010 18th International Conference on Geoinformatics, Jun. 2010, pp. 1–7.

[2] E. L. Ikanovic and A. Mollgaard, "An alternative approach to the limits of predictability in human mobility," EPJ Data Sci., vol. 6, no. 1, p. 12, Dec. 2017.

[3] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of Predictability in Human Mobility," Science, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.

[4] B. Cs. Csáji et al., "Exploring the mobility of mobile phone users," Physica A: Stat. Mech. Appl., vol. 392, no. 6, pp. 1459–1473, Mar. 2013.

[5] A. Cuttone, S. Lehmann, and M. C. González, "Understanding predictability and exploration in human mobility," EPJ Data Sci., vol. 7, no. 1, p. 2, Dec. 2018.

[6] International Telecommunication Union, "Measuring digital development. Facts and figures 2020," 2020. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf.

[7] R. Becker et al., "Human mobility characterization from cellular network data," Commun. ACM, vol. 56, no. 1, pp. 74–82, Jan. 2013.

[8] Y. Zhao, B. Y. Chen, F. Gao, and X. Zhu, "Dynamic community detection considering daily rhythms of human mobility," Travel Behav. Soc., vol. 31, pp. 209–222, Apr. 2023.

[9] S. Knezevic, "Evaluation of Human Mobility from Crowd-sourced Cellular Data," Ph.D. dissertation, Florida Inst. of Tech., Melbourne, FL, 2023. [Online]. Available: http://hdl.handle.net/11141/3699.

[10] A. Lind, A. Hadachi, and O. Batrashev, "A new approach for mobile positioning using the CDR data of cellular networks," in 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Jun. 2017, pp. 315–320.

[11] A. Sevtsuk and C. Ratti, "Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks," Journal of Urban Technology, vol. 17, no. 1, pp. 41-60, Apr. 2010.

[12] J. Pu, P. Xu, H. Qu, W. Cui, S. Liu, and L. Ni, “Visual analysis of people’s mobility pattern from mobile phone data,” in Proceedings of the 2011 Visual Information Communication - International Symposium (VINCI ’11), 2011, pp. 1–10.

[13] S. Isaacman et al., “Human mobility modeling at metropolitan scales,” in Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys ’12), 2012, p. 239.

[14] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González, “Unravelling daily human mobility motifs,” Journal of The Royal Society Interface, vol. 10, no. 84, p. 20130246, Jul. 2013.

[15] W. Sun, D. Miao, X. Qin, and G. Wei, “Characterizing User Mobility from the View of 4G Cellular Network,” in 2016 17th IEEE International Conference on Mobile Data Management (MDM), Jun. 2016, pp. 34–39.

[16] Z. Zhao, P. Zhang, H. Huang, and X. Zhang, “User mobility modeling based on mobile traffic data collected in real cellular networks,” in 2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS), Dec. 2017, pp. 1–6.

[17] S. Danafar, M. Piorkowski, and K. Krysczcuk, “Bayesian framework for mobility pattern discovery using mobile network events,” in 2017 25th European Signal Processing Conference (EUSIPCO), Aug. 2017, pp. 1070–1074.

[18] S. Danafar, M. Piorkowski, and K. Kryszczuk, “Bayesian framework for mobility pattern discovery using mobile network events,” in 2017 25th European Signal Processing Conference (EUSIPCO), Aug. 2017, pp. 1070–1074.

[19] S. Knezevic, I. Kostanic, and Z. Matloub, "Fundamental Spatial Parameters of Human Mobility," presented in The 2023 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE’23), 2023.

[20] Z. Matloub, I. Kostanic, and S. Knezevic, "Evaluation of Sampling Thresholds for Fundamental Spatial Parameters of Human Mobility," presented in The 2023 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE’23), 2023.

[21] R. A. Becker et al., "A Tale of One City: Using Cellular Network Data for Urban Planning," IEEE Pervasive Computing, vol. 10, no. 4, pp. 18–26, Apr. 2011.

[22] . Simini, M. C. González, A. Maritan, and A.-L. Barabási, “A universal model for mobility and migration patterns,” Nature, vol. 484, no. 7392, pp. 96–100, Apr. 2012.

[23] M. Dash et al., “CDR-To-MoVis: Developing a Mobility Visualization System from CDR data,” 2015 IEEE 31st International Conference on Data Engineering, pp. 1452–1455, Apr. 2015.

[24] J. Yang, Y. Shi, C. Yu, and S.-J. Cao, “Challenges of using mobile phone signalling data to estimate urban population density: Towards smart cities and sustainable urban development,” Indoor and Built Environment, vol. 29, no. 2, pp. 147–150, Feb. 2020.

[25] Y. Guo, J. Zhang, and Y. Zhang, “A method of traffic congestion state detection based on mobile big data,” 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(, pp. 489–493, Mar. 2017.

[26] F. Xiang, L. Tu, and B. Huang, “Inferring Barriers of Urban City Using Mobile Phone Record,” 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp. 850–855, Aug. 2013.

[27] P. Yang, T. Zhu, X. Wan, and X. Wang, “Identifying Significant Places Using Multi-day Call Detail Records,” 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, pp. 360–366, Nov. 2014.

[28] Z. Liu, Y. Qiao, S. Tao, W. Lin, and J. Yang, “Analyzing human mobility and social relationships from cellular network data,” 2017 13th International Conference on Network and Service Management (CNSM), pp. 1–6, Nov. 2017.

[29] L. Tongsinoot and V. Muangsin, “Exploring Home and Work Locations in a City from Mobile Phone Data,” 2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 123–129, Dec. 2017.

[30] V. Frias-Martinez, C. Soguero, and E. Frias-Martinez, “Estimation of urban commuting patterns using cellphone network data,” Proceedings of the ACM SIGKDD International Workshop on Urban Computing - UrbComp ’12, p. 9, 2012.

[31] Md. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, “Development of origin–destination matrices using mobile phone call data,” Transportation Research Part C: Emerging Technologies, vol. 40, pp. 63–74, Mar. 2014.

[32] L. Alexander, S. Jiang, M. Murga, and M. C. González, “Origin–destination trips by purpose and time of day inferred from mobile phone data,” Transportation Research Part C: Emerging Technologies, vol. 58, pp. 240–250, Sep. 2015.

[33] D. M. Bhandari, A. Witayangkurn, R. Shibasaki, and Md. M. Rahman, “Estimation of Origin-Destination using Mobile Phone Call Data: A Case Study of Greater Dhaka, Bangladesh,” 2018 Thirteenth International Conference on Knowledge, Information and Creativity Support Systems (KICSS), pp. 1–7, Nov. 2018.

[34] Q. Wang and J. E. Taylor, “Patterns and Limitations of Urban Human Mobility Resilience under the Influence of Multiple Types of Natural Disaster,” PLOS ONE, vol. 11, no. 1, p. e0147299, Jan. 2016.

[35] M. Pourmoradnasseri, K. Khoshkhah, A. Lind, and A. Hadachi, “OD-Matrix Extraction based on Trajectory Reconstruction from Mobile Data,” 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 1–8, Oct. 2019.

[36] S. Jiang, J. Ferreira, and M. C. González, “Clustering daily patterns of human activities in the city,” Data Mining and Knowledge Discovery, vol. 25, no. 3, pp. 478–510, Nov. 2012.

[37] E. Thuillier, L. Moalic, S. Lamrous, and A. Caminada, “Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data,” IEEE Transactions on Mobile Computing, vol. 17, no. 4, pp. 817–830, Apr. 2018.

[38] F. Calabrese, F. C. Pereira, G. di Lorenzo, L. Liu, and C. Ratti, “The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events,” Pervasive Computing. Pervasive 2010. Lecture Notes in Computer Science, vol. 6030, pp. 22–37, 2010.

[39] V. A. Traag, A. Browet, F. Calabrese, and F. Morlot, “Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference,” 2011 IEEE Third Int’l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int’l Conference on Social Computing, pp. 625–628, Oct. 2011.

[40] N. B. Ponieman, A. Salles, and C. Sarraute, “Human mobility and predictability enriched by social phenomena information,” Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1331–1336, Aug. 2013.

[41] K. Hajdú-Szücs and A. Kiss, “Event Detection on Call Detail Records,” Proceedings of the 10th International Conference on Applied Informatics, pp. 121–128, 2018.

[42] G. Pinter, L. Nadai, and I. Felde, “Analysis of Mobility Patterns During a Large Social Event,” 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000339–000344, Sep. 2018.

[43] N. E. Williams, T. A. Thomas, M. Dunbar, N. Eagle, and A. Dobra, “Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data,” PLOS ONE, vol. 10, no. 7, p. e0133630, Jul. 2015.

[44] L. Shi, W. Wang, W. Cai, Z. Wang, S. Zhang, and W. Zhou, “Mobility patterns analysis of Beijing residents based on call detail records,” 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, Oct. 2017.

[45] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, “Population flow drives spatio-temporal distribution of COVID-19 in China,” Nature, vol. 582, no. 7812, pp. 389–394, Jun. 2020.

[46] X. Lu, L. Bengtsson, and P. Holme, “Predictability of population displacement after the 2010 Haiti earthquake,” Proceedings of the National Academy of Sciences, vol. 109, no. 29, pp. 11576–11581, Jul. 2012.

[47] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, “Prediction of human emergency behavior and their mobility following large-scale disaster,” Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 5–14, Aug. 2014.

[48] X. Lu et al., “Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh,” Global Environmental Change, vol. 38, pp. 1–7, May 2016.

[49] E. Frias-Martinez, G. Williamson, and V. Frias-Martinez, “An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information,” 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 57–64, Oct. 2011.

[50] L. Bengtsson et al., “Using Mobile Phone Data to Predict the Spatial Spread of Cholera,” Scientific Reports, vol. 5, no. 1, p. 8923, Aug. 2015.

[51] H. Fang, L. Wang, and Y. Yang, “Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China,” Journal of Public Economics, vol. 191, p. 104272, Nov. 2020.

[52] N. Ayan et al., “Poster: Understanding Human Mobility during COVID-19 using Cellular Network Traffic,” 2021 IFIP Networking Conference (IFIP Networking), pp. 1–3, Jun. 2021.

[53] Alumni Ventures Group, "Change the World Webinar Series: Leveraging Location Data for Good - Castor Ventures Conversation with X-Mode CEO, Joshua Anton," retrieved Dec. 7, 2020.

[54] J. Reardon, "Proximity Tracing in an Ecosystem of Surveillance Capitalism – The AppCensus Blog," retrieved Dec. 7, 2020.

[55] Greater Houston Partnership, “Houston Metropolitan Statistical Area Profile,” July 29, 2021. https://www.houston.org/houston-data/houston-metropolitan-statistical-area-profile (accessed February 24, 2023).

[56] U.S. Census Bureau, “New York-Newark-Jersey City, NY-NJ-PA Metro Area,” 2021. Available : https://censusreporter.org/profiles/31000US35620-new-york-newark-jersey-city-ny-nj-pa-metro-area/ (accessed May 15, 2023).

[57] I. Brodsky, “H3: Uber’s Hexagonal Hierarchical Spatial Index,” 2018. https://eng.uber.com/h3/#:~:text=With%20this%20in%20mind%2C%20Uber,on%20a%20city%2Dwide%20level. (accessed Apr. 27, 2022).

[58] United States Census Bureau, “Metropolitan and Micropolitan Statistical Areas Population Totals: 2020-2022,” 2023. https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-and-micro-statistical-areas.html#v2022 (accessed Aug. 11, 2023).

[59] Apache Spark, “Unified engine for large-scale data analytics,” 2018. https://spark.apache.org/ (accessed Apr. 27, 2022).

[60] S. Hoteit, S. Secci, S. Sobolevsky, G. Pujolle, and C. Ratti, “Estimating Real Human Trajectories through Mobile Phone Data,” in 2013 IEEE 14th International Conference on Mobile Data Management, IEEE, Jun. 2013, pp. 148–153. doi: 10.1109/MDM.2013.85.

[61] [20] J. Smart, W. Powell, and S. Schey, “Extended range electric vehicle driving and charging behavior observed early in the EV project”, SAE Technical Paper. 2013.

[62] D. Ashbrook and T. Starner, "Learning significant locations and predicting user movement with GPS," Proceedings. Sixth International Symposium on Wearable Computers,, Seattle, WA, USA, 2002, pp. 101-108

[63] K. Yadav, A. Kumar, A. Bharati and V. Naik, "Characterizing mobility patterns of people in developing countries using their mobile phone data," 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 2014.

[64] T. M. T. Do and D. Gatica-Perez, "The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data," IEEE Trans. Mobile Comput., vol. 13, no. 3, pp. 638–648, 2014.

[65] Z. Cheng, J. Caverlee, K. Lee and D. Z. Sui, "Exploring millions of footprints in location sharing services", ICWSM, vol. 2011, pp. 81-88, 2011.

[66] S. Isaacman et al., "Ranges of human mobility in Los Angeles and New York," 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011, pp. 88-93

[67] U.S. Department of Transportation - Federal Highway Administration, "Highway Statistics 2019," [Online]. Available: https://www.fhwa.dot.gov/policyinformation/statistics/2019/. [Accessed: 27-Sep-2023].

[68] How Much Time Do Americans Spend in Their Cars Every Year?," Space Coast Daily, July 2020. [Online]. Available: https://spacecoastdaily.com/2020/07/how-much-time-do-americans-spend-in-their-cars-every-year/

[69] Americans spend 19 full work days a year stuck in traffic on their commute," New York Post, 19-Apr-2019. [Online]. Available: https://nypost.com/2019/04/19/americans-spend-19-full-work-days-a-year-stuck-in-traffic-on-their-commute/. [Accessed: 27-Sep.-2023].

[70] "Study: Americans spend 18 days in their car per year, forge close bonds with a vehicle," The Car Connection, [Online]. Available: https://www.thecarconnection.com/news/1122782_study-americans-spend-18-days-in-their-car-per-year-forge-close-bonds-with-a-vehicle. [Accessed: 27-Sep.-2023].

[71] M. Papandrea, K. K. Jahromi, M. Zignani, S. Gaito, S. Giordano, and G. P. Rossi, "On the properties of human mobility," Computer Communications, vol. 87, pp. 19-36, 2016

[72] V. Hedrih and A. Hedrih, “Chapter 4: Deviation from normal distribution,” Interpreting Statistics for Beginners: A Guide for Behavioural and Social Scientists, 2022, p. 77.

[73] J.L. Horn “A rationale and test for the number of factors in factor analysis” Psychometrika 30,179–185(1965).

[74] T. W. Anderson and H. Rubin, "Statistical Inference in," in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California, Dec. 1954, Jul.-Aug. 1955, p. 111.

[75] C. E. Bonferroni, "Il calcolo delle assicurazioni su gruppi di teste," in Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy, 1935, pp. 13-60.

[76] J. Cohen, Statistical Power Analysis for the Behavioral Sciences. 2013.