FUNDAMENTAL UNDERSTANDING ON THE USE OF MOBILE PHONE DATA FOR TRANSPORT APPLICATIONS

Matsumi Suzudo
BEng

Submitted in fulfilment of the requirements for the degree of
Master of Philosophy

School of Civil and Environmental Engineering
Faculty of Engineering
Queensland University of Technology
2021

Keywords

Call Detail Records (CDR), Geographical Analysis, Mobile Phone Data, Origin-Destination Matrix, Cellular Signalling Data, Sightings Data, Location Area Update, Handover, Geographical Grouping, Identification of Meaningful Locations, Land Use, Population Distribution, Travel Time, Travel Speed, Traffic Density, Travel Demand, Purpose-specific OD Estimation, Mode-specific OD Estimation, Synthetic CDR Data

Abstract

Among various types of new sources of data which are utilized for mobility studies, data extracted from mobile phone play an important role. This is because mobile phone data can represent the movements of individuals well, and the attempts by researchers are continuing. Thus, an updated review for the understanding of the data and the applications has been conducted, so that the current achievements and limitations are revealed.

Regarding the review of mobile phone data, there have been various types of mobile phone data, which can be classified into data for billing purposes and management purposes. Whichever type of data, they can represent the movements of each user well, but some bias emerging from various types of users still needs to be considered.

Regarding the applications of the data, a review has been conducted in the fields of mobility studies and other related fields. Geographical analysis, which includes geographical grouping, population distribution, identification of meaningful locations, and estimation of land use, are some of the fields that researchers have managed to utilize mobile phone data. The reliability for the identification of meaningful locations has been supported by a simple experiment. Meanwhile, there have been applications in mobility studies, and some have been used in practice by using mobile phone data for management purposes. Estimation of travel demand is the field that there has been a significant amount of research. There could be seen that the reliability of Origin-Destination (OD) matrices with trip purposes is high, but the estimation of modes remains to be a challenge.

This research has brought an understanding of mobile phone data and its applications. It has also revealed the fields where mobile phone data works well for estimation, in which further applications can be done in the future, such as mode-specific OD matrices.

Table of Contents

Keywords i

Abstract ii

Table of Contents iii

List of Figures v

List of Tables vi

List of Publications i

List of Abbreviations i

Statement of Original Authorship ii

Acknowledgement iii

Chapter 1: Introduction 1

1.1 Background 1

1.2 Research Questions 2

1.3 Objective 2

1.4 Task 2

1.5 Scope 3

1.6 Thesis Outline 3

Chapter 2: Fundamental Understanding on the Data from Mobile Phone Network 5

2.1 Background 5

2.2 System of Mobile Phone Network 5

2.3 Data Extracted from Mobile Phone Network 10

2.4 Discussion 13

2.5 Summary 13

Chapter 3: CDR Data for Transport and related Applications 15

3.1 Background 15

3.2 Literature Review 15

3.3 Application on Geographical Analysis 18

3.4 Application on Transportation 21

3.5 Application on Travel Demand 22

3.6 Summary 24

Chapter 4: Synthetic CDR and Experiments using Real Data 27

4.1 Background 27

4.2 Methodology 27

4.3 Experiment 30

4.4 Discussion from the results 33
4.5 Application of the Synthetic CDR 33
Chapter 5: Conclusion 34
Bibliography 37

List of Figures

Figure 1: Network Architecture of CCS 5

Figure 2: Comparison of different types of cells 7

Figure 3: Locations of Base Stations in the Brisbane CBD Area (retrieved from RFNSA Website) 8

Figure 4: Locations of Base Stations in the southern suburbs of Brisbane (retrieved from RFNSA Website) 8

Figure 5: Locations of Base Stations around Jondaryan, QLD (retrieved from RFNSA Website) 9

Figure 6: Concept of LAU 10

Figure 7: Clustered Communities based on CDR in Belgium (Adam et al., 2017) 18

Figure 8: The regions identified from modularity maximisation represented by different colours, and the strongest 80% links between each local grid. (Ratti et al., 2010) 19

Figure 9: Clustering result for land use types in Singapore (circled area being the misclassified area) (Pei et al., 2014) 21

Figure 10: Generation framework of the synthetic CDR 28

Figure 11: Description of the concept of overlaying the connection events with trajectory 28

Figure 12: Trajectory of one day collected by participant E, cell tower locations of Vodafone Australia provided from RFNSA Website and estimated locations of connection events 29

Figure 13: The number of connections in each of the candidate cells shown in Table 10 31

List of Tables

Table 1: Types of cells (source: Ghadialy (2017)) 6
Table 2: Example of Raw CDR (Houlton, 2011) 11
Table 3: Example of triangulated records 12
Table 4: Example of LAU Register data (Calabrese, 2011a) 13
Table 5: List of review papers with the explored fields and the set of data 16
Table 6: List of Studies Mentioned in Section 3.4 22
Table 7: List of papers mentioned in Section 3.5 26
Table 8: Part of Synthetic CDR relevant to trajectory shown in Figure 12 30
Table 9: Details of each Estimated CDRs 30
Table 10: List of Individuals' Meaningful Locations 31

List of Publications

CONFERENCES

1. **Matsumi Suzudo,** Krishna N. S. Behara, Ashish Bhaskar, Insights on the mobile phone data applications, In World Conference on Transport Research - WCTR 2019, 26-31 May 2019, Mumbai, India.

List of Abbreviations

BS	Base Station
BSC	Base Station Controller
CBD	Central Business District
CCS	Cellular Communication System
CDR	Call Detail Records
CNS	Core Network Subsystem
CSD	Cellular Signalling Data
GPS	Global Positioning System
IPDR	Internet Protocol Detail Record
LA	Location Area
LA-ID	Location Area ID
LAU	Location Area Update
LTE	Long Term Evolution
OD	Origin-Destinations
QLD	Queensland
RA	Routing Area
RFNSA	Radio Frequency National Site Archive
SMS	Short Messaging Services
TA	Tracking Area
UMTS	Universal Mobile Telecommunications System

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: QUT Verified Signature

Date: 16/07/2021

Acknowledgement

I express my gratitude to the following people.

Firstly, I would like to send my greatest gratitude to Aspro Ashish Bhaskar and Dr Krishna N. S. Behara. I believe that this journey has been impossible without having them as my supervisor and mentors. Aspro Ashish Bhaskar, my principal supervisor, welcomed me with open hands when I first made a contact back in 2017, and when I joined the team in April 2018. He had been caring about myself ever since, which made my journey much easier and meaningful. Dr Krishna Behara has been my mentor, later my associate supervisor, and he was helping with my work constantly. I was lucky enough for having them as my supervisors, and would like to thank Prof Masao Kuwahara in Tohoku University, Japan for giving me a pathway to come here. I would also like to thank Prof James McGree, my associate supervisor, for helping me with my research by giving comments.

I would also like to send my gratitude to my parents for supporting me both financially and mentally.

Finally, I thank all my fellows who I came to know in QUT throughout this journey. They have made my 2 years here more lively and joyful, and this friendship are one of my most important ones.

Chapter 1: Introduction

This chapter outlines the background (Section 1.1), research questions (Section 1.2), and the objectives (Section 1.3). Section 1.4 describes the scope of this research and provides definitions of terms used, and Section 1.5 provides the task. Finally, Section 1.6 provides an outline of the remaining chapters of the thesis.

1.1 BACKGROUND

The utilization of a new source of data into mobility studies has been a hot topic with the diffusion of various devices. Some new sources of data include ones from mobile phones, Smart Cards (SC), road tolling systems, In-vehicle GPS, and Bluetooth (BITRE, 2014). For example, SC data, which are used for collecting fare automatically on public transports, gives rich information on understanding a multiday travel regularity of public transport passengers. Specifically, it can be applied to understand passenger behaviour, passenger segment, and trip purpose (Kieu et al., 2015); and transit demand between different Origin-Destination (OD) pairs (Hussain, Behara, et al., 2021; Hussain, Bhaskar, et al., 2021a, 2021b). For Bluetooth data, the travel activity of vehicles that carry BT-equipped devices can be extracted by using Bluetooth Media Access Control Scanner (BMS) which are attached on roadsides. The examples of the applications of Bluetooth data include travel time estimation (Bhaskar & Chung, 2013), development of vehicle trajectories (Advani et al., 2021; Michau et al., 2017), comparison of travel patterns (Behara et al., 2018; Behara et al., 2020a, 2020b; Hussain, Behara, et al., 2021), and OD estimation (Behara et al., 2020c, 2021; Michau et al., 2017).

Regarding mobile phones, the breakthrough since the 1990s has enabled a significant number of people to use this device, and as of 2019 as many as 53% of the people living in the world use mobile phones (ITU, 2019). In Australia, 96% of adults currently own at least one mobile phone (ACMA, 2019). The set of data that are collected for billing and management purposes can contain a rich amount of information related to the users' daily behaviours, and many researchers have been analysing this data. Their applications include estimation of traffic volumes (Caceres et al., 2007), travel speeds (Jaume & Montero, 2015), travel times (Barcelö et al., 2010), and travel demand (Calabrese et al., 2011). This is continuing up until these past few years, and a review of the current achievements shall contribute to organising the remaining challenges for further applications.

As a research gap, it is recognized that a critical review is needed for recognizing the challenges. This has brought to the motivation to conduct a review of the mobile phone data, its features, and applications, giving insights on its current limitations and challenges.

1.2 RESEARCH QUESTIONS

Recognizing the research gap mentioned in the previous section, this research is planned and carried out to answer the following questions:

1. What is the state-of-the-art for using mobile phone data for mobility studies?
2. What are the limitations on the current state of the art?

1.3 OBJECTIVE

This research aims at revealing some unaddressed challenges of mobile phone data on mobility studies, suggesting a pathway for further applications. Specifically, the objectives are listed below:

• Provide a fundamental understanding of the mobile phone data.

• Understand the current limitations of mobile phone data in transport applications. This is based on a detailed literature review.

1.4 TASK

Based on the objectives proposed, the following are set to be the tasks.

- Conduct a literature review on the fundamentals of mobile phone data.
- Conduct a literature review on the application of mobile phone data in the field of transportation and related fields.
- Holding experiments using the actual mobile phone data for its understanding based on the literature reviews.

1.5 SCOPE

This research is focused on the applications of mobile phone data for mobility studies and other closely related fields including geographical analysis. Applications of the other fields are not considered.

The mobile phone data is defined as the set of data that is passively generated from the mobile phone network, and it does not include that are collected intensively, such as GPS data or additional phone apps for data collection.

1.6 THESIS OUTLINE

The chapters of this chapter can be described as the following.

Chapter 1 (this chapter) is an introductory chapter that describes the background, research questions, aim, objectives, and scope of this research.

Chapter 2 provides a fundamental understanding of mobile phone data and its applications, revealing the limitations and challenges.

Chapter 3 reviews the literature on the use of CDR data for transport and other related applications.

Chapter 4, describes the actual CDR data with some experiments conducted, based on the literature review.

Chapter 5 includes the summary and conclusion of this research.

Chapter 2: Fundamental Understanding on the Data from Mobile Phone Network

2.1 BACKGROUND

This chapter describes the fundamentals of the mobile phone network for the understanding of how mobile phones system work, and what kind of data can be extracted. In section 2.2, the system of the mobile phone network is described, along with some important features. In Section 2.3, the types of data passively extracted from the mobile phone network are described. Thereafter, section 2.4 is a discussion of the features of the sets of data, and finally the chapter is summarised in Section 2.6.

2.2 SYSTEM OF MOBILE PHONE NETWORK

The system used for a mobile phone network is typically called a cellular communication system (CCS), which is designed to provide communication for each mobile device. The service area is geographically divided into small regions, called cells, which is the range of a service antenna. Each antenna is powered by a station called the base station (BS), and one BS controls one or more directional antennas attached to a tower or an existing building.

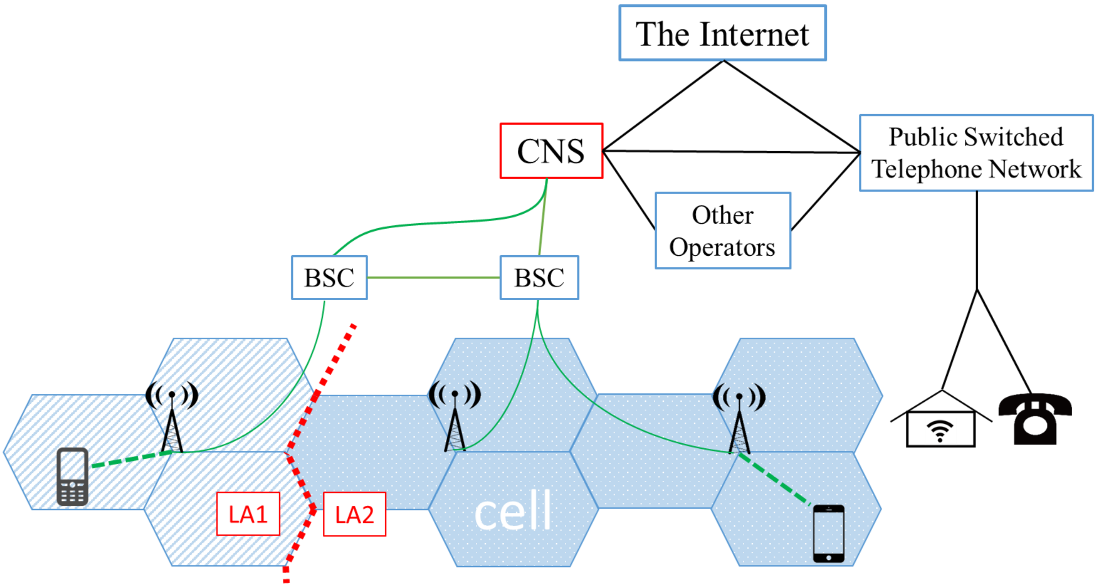


Figure 1: Network Architecture of CCS

Figure 1 describes the network architecture of CCS. The BS is connected to the Base Station Controller (BSC) with a landline or with a microwave in limited cases, and it acts as a mediator between the BS and the Core Network Subsystem (CNS). The Core Network Subsystem (CNS) is a subsystem that is in charge of the connections with mobile phones, fixed-line phones, phone network of other operators', and the Internet, managing all the connections and communication records (Ghadialy, 2017).

2.2.1 Types of Cells

The size of each cell differs depending on its location and purpose as shown in Table 1 (Ghadialy, 2017) and Figure 2. Table 1 describes the coverage distance, the common height of the antenna, the capacity of the users connected at once, common locations, and the relative cost of each cell. Macro-cells are the large cells often set in rural areas, which cover a large number of users in a wide area. Macro-cells often require an installation of tall antennas as such they are more expensive compared to other cells. Meadow-cells are often placed in rural towns where it is not covered well with a macro-cell. With the low frequencies of macro-cells, it is difficult for the signals to cover indoors and shadows of mountains, and meadow-cells play the role to cover these areas. Micro-cells are generally used for urban areas, where the demand for communication is higher as compared to that of rural areas. These cells can cover a large number of people within an urban zone. Pico-cells are often used to meet the requirements where there is a high density of usages such as shopping malls and airports. They are usually the least expensive as compared to other types and are also used inside tunnels, trains, airplanes, ferries, and mountain huts which a connection is required in a small range outside of the service area.

Table 1: Types of cells (source: Ghadialy (2017))

	Coverage Distance	Height of antenna (m)	Capacity of Users	Locations	Cost
Macro-cells	25-40km	15-25	50-200	Rural	High
	1-10km	5-15	32-64	Rural Towns	Medium
Micro-cells	500m-3km	8-10	32-200	Urban	Medium
Picocells	<250m	N/A	32-64	Populated Indoors	Low

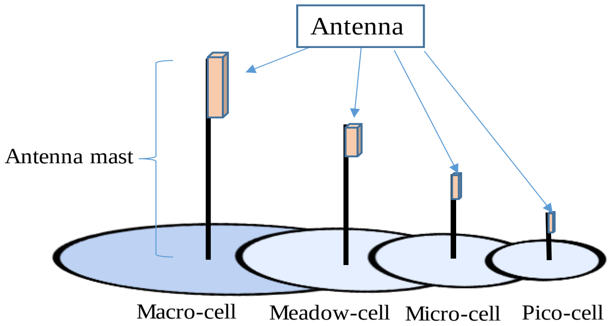


Figure 2: Comparison of different types of cells

As an example of the distributions of the cells, Figure 3, Figure 4, and Figure 5 respectively show the distribution of the base stations in the Brisbane Centre Business District (CBD) area, the Southern Suburbs of Brisbane, and a rural area around Jondaryan, QLD. These images are retrieved from the Radio Frequency National Site Archive (RFNSA) website (https://www.rfnsa.com.au/). The ID unique to each base station is shown on each one. The Brisbane CBD area is typically where microcells are implemented, and several base stations by multiple operators can be found in a range of a few hundred metres. Meanwhile, the area of the Southern Suburbs of Brisbane is where meadow-cells are commonly used, and the area can be fully covered with each Base Station covering its few kilometres radius. On the other hand, in the surroundings of Jondaryan, QLD is where macro-cells are found, and a single base station covers tens of kilometres radius. In summary, the coverage of a single base station and the size of a cell differs greatly depending on its geographical conditions.

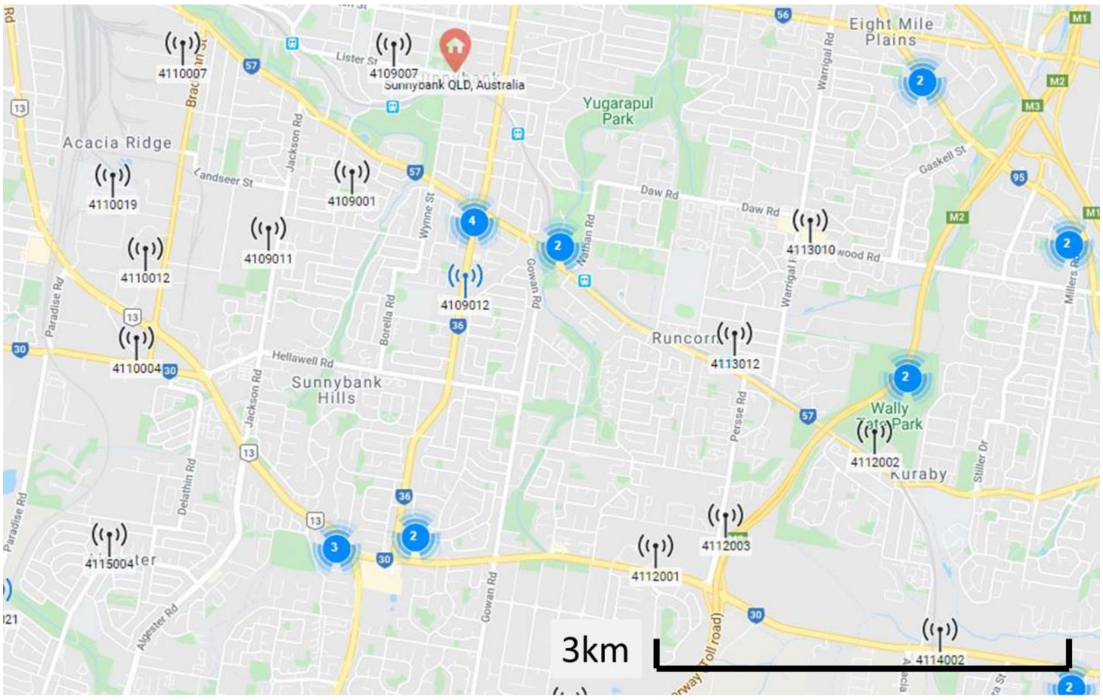


Figure 3: Locations of Base Stations in the Brisbane CBD Area (retrieved from RFNSA Website)

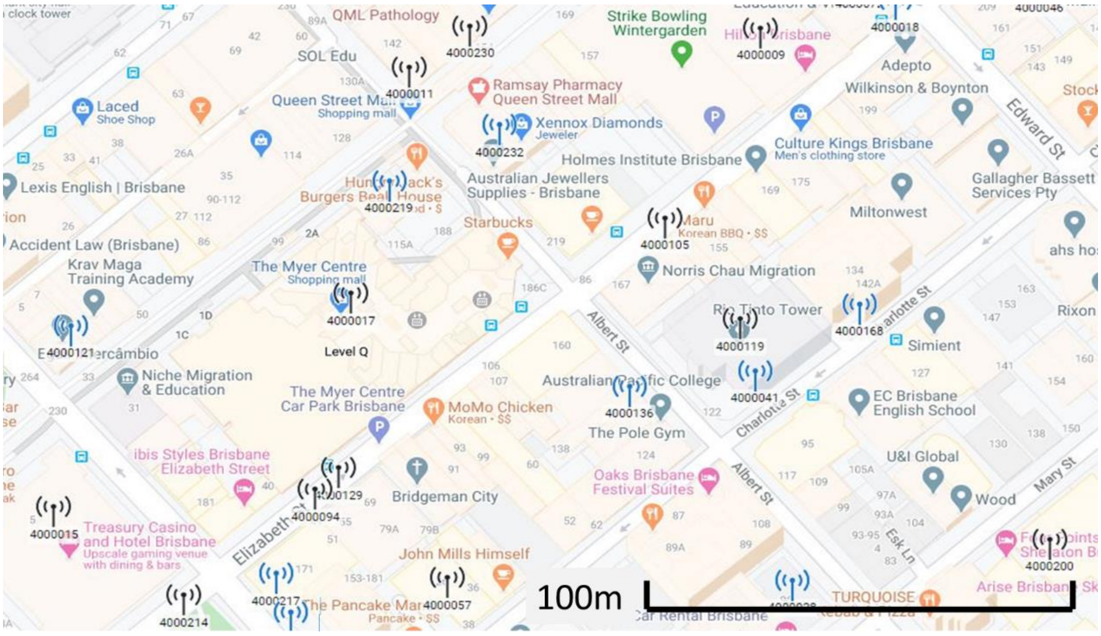


Figure 4: Locations of Base Stations in the southern suburbs of Brisbane (retrieved from RFNSA Website)

The blue symbols with numbers indicate multiple Base Stations at the location

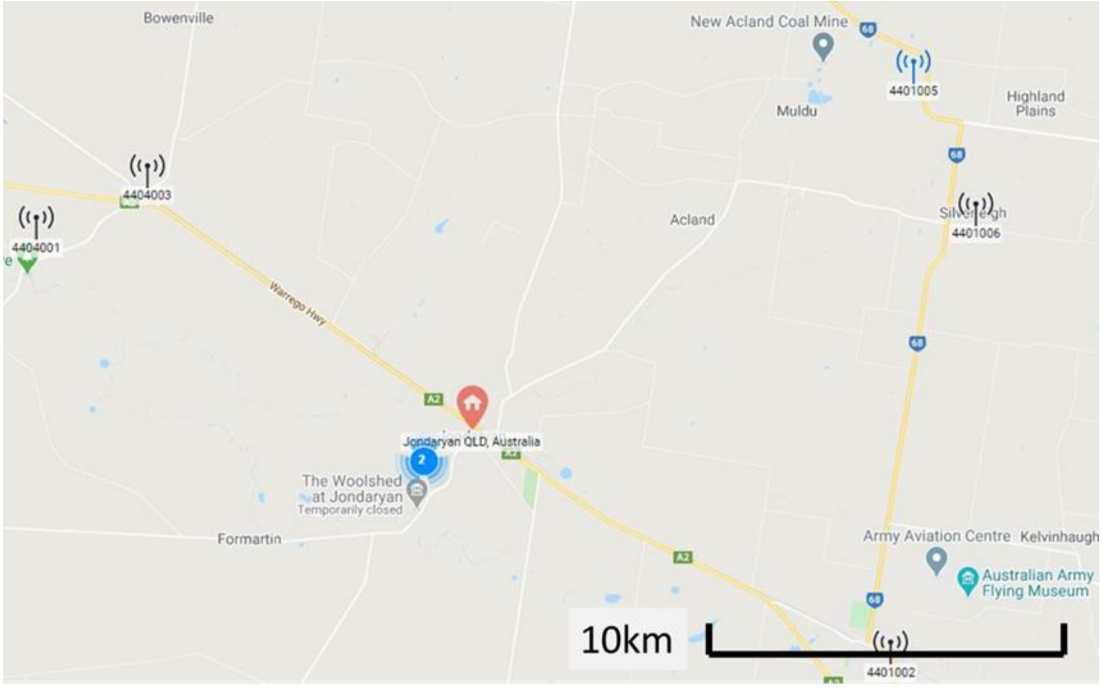


Figure 5: Locations of Base Stations around Jondaryan, QLD (retrieved from RFNSA Website)

The blue symbols with numbers indicate multiple Base Stations at the location

2.2.2 Location Area Update

One system to note in the CCS is the Location Area Update (LAU) system (Figure 6). Location Area (LA) is a set of base stations, dozens and hundreds, grouped for optimizing signals. This LA is called “Routing Area (RA)” for Universal Mobile Telecommunications System (UMTS) (3G Standard network) and “tracking area (TA)” for Long Term Evolution (LTE) (4G standard for communication). To track the phone, each device’s current location area ID (LA-ID) is recorded in the CNS. When the CNS starts a connection (call and/or text) with the phone, a signal that notifies the current LA-ID is sent from all of the BSs in the recorded LA. When a mobile phone detects that it is in a new LA, a message will be sent to the CNS so that the current LA will be updated, and this process is called a Location Area Update (LAU). If the LA is too large, the same signals would be sent from too many BSs to call a device, and if it is too small, LAU happens too often. Thus, the size of LA is the optimum of these constraints, which makes the connections more effective (Telebeansolutions, 2015).

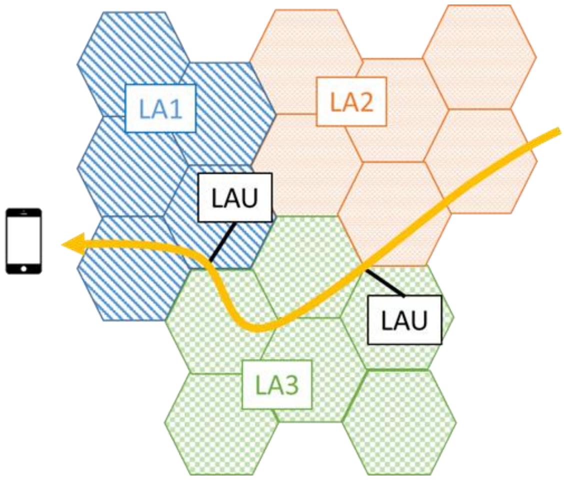


Figure 6: Concept of LAU

2.3 DATA EXTRACTED FROM MOBILE PHONE NETWORK

The data that can be extracted from the mobile phone network is a result of location updates (not to be confused with location area update (LAU)), which can be classified into event-triggered updates and network-triggered updates. The event-triggered updates occur when a mobile device:

- Receives or makes a call,
- Receives or makes an SMS,
- Is connected to the Internet,

The set of event-extracted data is usually either of the following:

- Call Detail Record (CDR)
- Internet Protocol Detail Record (IPDR)

On the other hand, the event-triggered updates are recorded when a mobile device:

- Is switched on and gets connected.
- Moves into a new cell (handover)
- Makes a Location Area Update
- Makes a periodic update

The set of network-extracted data are often known as Cellular Signalling Data or Sightings Data.

2.3.1 Call Detail Record (CDR)

CDR is a set of event-triggered data created by the mobile phone operators for billing and management purposes, which records call events and text events, and often the Internet usage as well. A call event is when the phone receives or sends a call, text event is when the phone sends or receives an SMS. The cell ID of the device and its timestamp is recorded, and this has triggered many researchers to use this for analysis. The raw CDR format is shown in Table 2. Here, “Begin” and “End” respectively shows when a connection event began and ended. “Type” shows the type of connections, such as call, SMS, or Internet connection (described here as GPRS). “In/outward” shows whether this connection event occurred by the action of the user’s phone, or by another device. “Latitude” and “Longitude” show the location where the BS is located, and “Direction” shows the direction of the antenna is facing. Finally, “Cell-Id_A” shows the ID of the unique to each cell (Houlton, 2011).

Table 2: Example of Raw CDR (Houlton, 2011)

Begin	End	Type	In/outward	Longitude	Latitude	Direction	Cell-Id_A
8/31/09 7:57	8/31/09 8:09	GPRS	Outward	13.39611	52.52944	30	45830
8/31/09 8:09	8/31/09 8:09	GPRS	Outward	13.38361	52.53	240	59015
8/31/09 8:09	8/31/09 8:15	GPRS	Outward	13.37472	52.53028	120	1845
8/31/09 8:15	8/31/09 8:39	GPRS	Outward	13.37472	52.53028	120	1845
8/31/09 8:20			Outward				

2.3.2 IP Detail Record (IPDR)

Similar to CDRs, IPDR is a set of event-triggered data that records the individual usage of the Internet, which records subscriber ID, type of website, timestamp, and the number of bytes transmitted that are primarily used for billing purposes (Calabrese, 2011b). Although it depends on the frequency of the record, as seen at call detail records, it can be expected that it could well represent the moving behaviour of the individual.

2.3.3 Sightings Data

Sightings data, also known as mobile sensing data or cellular signalling data (CSD), is a set of network-triggered data for management purposes. The basic concept of mobile phone triangulation is to measure the distance from base stations by the time-lapse of transmission and to locate the device. They are generated for operation and billing so that the connection is maintained, and billings are correct (Wang & Chen, 2018). Data are triggered on various occasions such as the following:

- Cell phone events (call, SMS, and/or data)
- Location Area Update (LAU)
- When the phone is switched on
- Periodic updates

The frequency of detections differs greatly between the operators, and each device can still be sighted periodically event without phone activity. The accuracy of the location is 300m on an average in urban areas, which is a higher resolution than tower-based estimation such as CDRs (Chen et al., 2014).

Table 3: Example of triangulated records

Order	1	2	3
Field Name	Mobile Device ID	Time	Location Estimate

As mentioned in Section 2.2.2, Location Area Update (LAU) is a system that is used for optimising the amount of transmission, and the register data of this is a set of network-triggered data for storing the LAUs. The movement between location areas can be detected, which makes it possible to track a mobile device. Table 4 shows an example of the LAU register data. LA1 shows the new LA after LAU, and C1 shows the cell ID where the device is located (Calabrese, 2011a). This would be identified as network-triggered data, and it can be included as part of sightings data. However, compared to sightings data, LAU is recorded each time when the phone enters a new area, as long as the phone is switched on. This means that it can have high accuracy on how the individual has moved between Location Areas.

Table 4: Example of LAU Register data (Calabrese, 2011a)

Order	1	2	3	4
Field Name	Time	Encrypted phone ID	LA1	C1

Meanwhile, handover is a process where the connected cell is transferred from one to another during a communication such as phone calls or data session without disconnection from the network. The handover data reflects the movement from one place to another if a call or a data session is occurring while the user is moving (Calabrese, 2011a), and can be classified as network-triggered data**.**

2.4 DISCUSSION

As described in the previous section, CDR and CSD are collected for different purposes from each other, and so the features that they can contain also differ. The occurrence of the timestamps for CDR heavily relies on the usage of individuals, as they are mostly recorded when the user is using the device. Meanwhile, this will not always be the case with sightings data, depending on the algorithm utilized by the operators for recording the data. Because the frequency of data depends on the individual usage, the accuracy of the estimated location also depends on it (Schlaich et al., 2010). This would mean that there can be biases that emerges from users with different usage of the device, and such biases needs to be considered.

2.5 SUMMARY

The Mobile Phone Network consists of the connection between the phones and devices, managed by the Core Network Subsystem. Each device is connected to a base station equipped with an antenna, and the range of each base station is called a cell. The data passively extracted from the network includes the data for billing purposes (CDR and IPDR) and data for management purposes (Sightings Data). From these data, it is possible to roughly estimate how each person usually travels. However, the accuracy of the estimation may contain some bias that emerges from the usage of each device.

Chapter 3: CDR Data for Transport and related Applications

3.1 BACKGROUND

The fact that mobile phone data is rich enough to extract mobility from large number of people has attracted attention of many researchers and analysts. In this chapter, a detailed review of studies on the applications of mobile phone data in transportation and related fields is presented along with a real-world experiment. Particularly, this chapter discusses up-to-date applications, limitations, and challenges of mobile phone data on origin-destination (OD) estimation. Rest of the chapter is organised in the following order. Section 3.2 presents the procedure adopted to conduct the literature review and compares the relevant studies. Section 3.3 discusses the mobile phone data application on geographical analysis. This includes geographical grouping, population distribution, identification of meaningful locations and estimation of land use. Section 3.4 is on the applications related to transport, except for travel demand and OD estimation. This includes traffic density, traffic counts, travel speed and trajectory estimation. In Section 3.5, various approaches on travel demand and OD estimation are presented and discussed, revealing the current limitations and challenges. Finally, Section 3.6 summarises this chapter.

3.2 LITERATURE REVIEW

As described in Table 5, there have previously been several reviews on the applications of mobile phone data. These have been searched on Web of Science and Scopus by the terms “mobile phone OR cell phone” and “review”, and the papers which are not review papers on the field of transportation and geographical analysis being filtered out. There were 102 papers found, and 10 review papers shown in Table 5 remained after the filtration.

Table 5: List of review papers with the explored fields and the set of data

Author	Title	Fields			Data		
		Geographical Analysis	Transportation	Other	CDR	CSD	Other
Caceres et al. (2008)	Review of traffic data estimations extracted from cellular networks	○	×	×	×	○	×
Steenbruggen et al. (2013)	Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities	○	○	×	×	○	×
Blondel et al. (2015)	A survey of results on mobile phone datasets analysis	○	×	×	○	○	○
Chandrasekar (2015)	Big data and transport modelling: opportunities and challenges	○	○	×	×	○	○
Rojas IV et al. (2016)	Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data	○	×	×	○	×	×
Anda et al. (2017)	Transport modelling in the age of big data	○	×	×	○	×	×
von Mörner (2017)	Application of call detail records-chances and obstacles	○	○	×	○	×	○
Wang et al. (2018)	Applying mobile phone data to travel behaviour research: A literature review	○	×	×	○	○	○
Huang et al. (2019)	Transport mode detection based on mobile phone network data: A systematic review	×	○	×	○	×	○
Simini et al. (2019)	Human Mobility from theory to practice: Data, Models and Applications	○	○	×	○	○	×

Caceres et al. (2008) one of the earliest works on the usage of Mobile Phone Data (MPD) to measure traffic characteristics, such as traffic counts, speed, travel time and traffic density and OD Matrix. This paper discussed the advantages and disadvantages of MPD in detail. Sightings was explored to discuss the applications, and CDR was not included in the review. Blondel et al. (2015) explored the applications beyond transportation, such as personal mobility, geographical partitioning, urban planning, and help towards development as well as security and privacy issues. Steenbruggen et al. (2013) reviewed the research and industry projects focused on applications of mobile phone data for mobility studies. Steenbruggen et al. (2013)'s study showcased how actual MPD can be utilized for extracting daily patters of mobility. Chandrasekar (2015) focused on the application of big data, including mobile phone data, Smart Card and smartphone apps, whether it can replace traditional methods for traffic modelling. It has been noted that CDR is helpful for OD-estimation, and combination with other source of data could make this more accurate. Rojas IV et al. (2016) reviewed the applications of MPD and GPS records, especially on the data processing. It implied that further research is required for the sample bias of the data, and combination with demographic data could help with this issue. The presentation by Simini et al. (2019) briefly described how human mobility, especially trajectories and moving patterns, can be extracted from "mobility data", such as CDR, and GPS data. Anda et al. (2017) reviewed the applications of various "new Big Data sources" such as mobile phone data and smartcard data on mobility analysis. von Mörner (2017) briefly reviewed the application of CDR in the field of transportation, revealing the remaining obstacles, such as the limitation on trip identification. Wang et al. (2018) reviewed the applications of MPD on individual travel behaviour, noting that further analysis can be done to extract hidden patterns of movement, as this is a new type of big data. Huang et al. (2019) reviewed on the estimation of travel modes from mobile phone data and GPS records. It was described that extracting travel modes from individual users remains challenged.

This literature review explores widely on transport applications and some related fields, updating the remaining issues and challenges noted in these existing literatures, suggesting insights for further applications. The papers have been searched by Google Scholar. This included searching for papers that cited the existing review papers as well as articles that conducted similar research.

3.3 APPLICATION ON GEOGRAPHICAL ANALYSIS

Here in this section, various applications of mobile phone data on geographical analysis are described.

3.3.1 Geographical Partitioning

The CDR data enables geographical analysis of human mobility due to its ability to record the timestamps of located cells. This means that extracting communities on geographical basis (geographical partitioning) and other geographical analysis is possible. Geographical grouping was a hot topic for research since the 1960s, often applying the multivariate algorithm. However, it was not until the 1990s that the geographical grouping through the complex network was developed. This was due to the development of computers and the field of social network science (Fukumoto & Okamoto, 2012).

Drawing the social borders between communities and regions by analysing the interactions recorded on CDR was one idea. These were often conducted using traditional community extraction methods such as modularity maximization. Adam et al. (2017) and Ratti et al. (2010) respectively analysed the frequency and duration of calls between communities and regions, revealing how the 'cultural barrier' exists between these optimal partitions. As a result, the 'social borders' were developed through CDRs as shown in Figure 7 and Figure 8. Similarly, Chi et al. (2016) further seized multiple levels of communities, by implementing not only the call frequencies and durations of calls but also the geographical density and distributions of cells. The

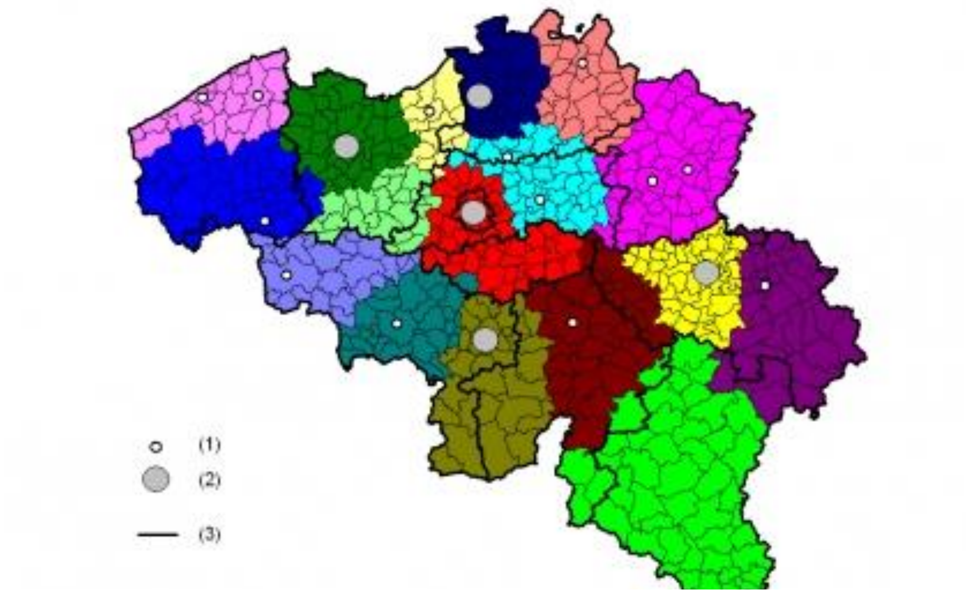


Figure 7: Clustered Communities based on CDR in Belgium (Adam et al., 2017)

findings of these studies were based on geographical information at a "cell" level and the penetration rate of CDR. From these results, it can be said that a picture of modality of communities can be seized from the analysis of CDR.

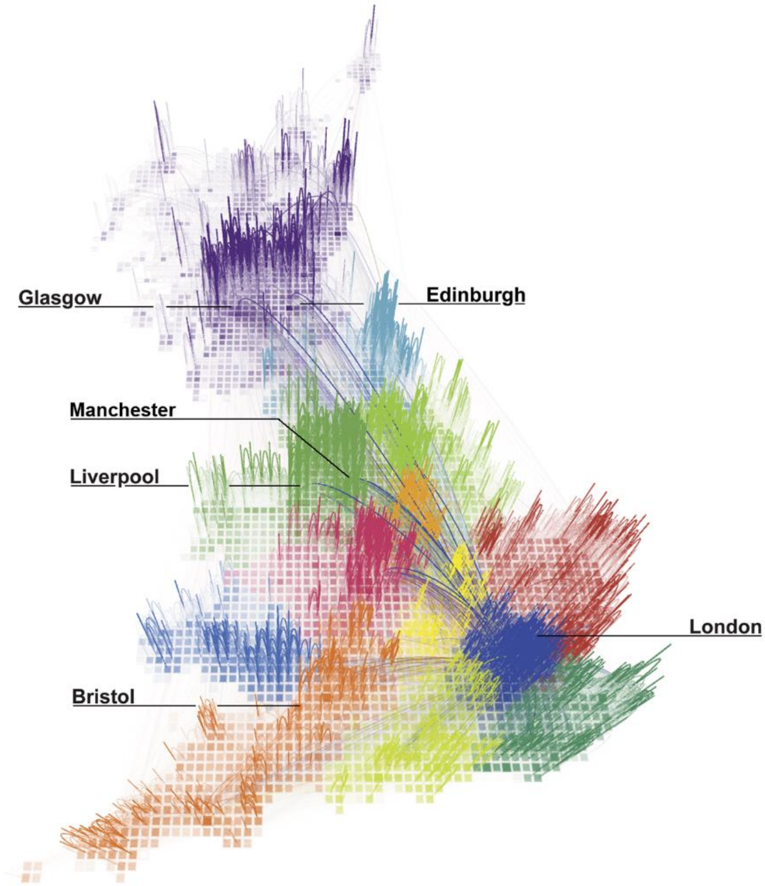


Figure 8: The regions identified from modularity maximisation represented by different colours, and the strongest 80% links between each local grid. (Ratti et al., 2010)

3.3.2 Population Distribution

CDR was applied to estimate the population in different areas, including time-dependent population distributions. Time-dependent population distributions can be estimated on a daily (day and night), weekly (weekdays and weekends), seasonally, on an annual basis. Deville et al. (2014) estimated the call-density; that is, the number of calls generated in each cell in real-time, and revealed a good correlation between weekly and daily distributions compared to the census data. However, this correlation was not as good in developing countries (e.g., DR Congo) as it was in developed countries (e.g., Portugal and France). Other applications include the detection of residential locations (as by the International Telecommunication Union (2017)). These estimations using CDR can be an alternative to the traditional estimation by the census in the point that estimation can be conducted quickly, and they can also be helpful where conducting the census is an expensive process because of poor transportation infrastructure, mainly in developing countries.

3.3.3 Identification of Meaningful Locations

Identifying geographical locations of home and work has been another application of CDR data. Ahas et al. (2010) and Isaacman et al. (2011) developed models to detect home and work locations and validated them with the Census data. These studies showed that home and work locations can be detected from analysing frequently visited locations and the times of day of those stays.

3.3.4 Estimation of Land Use

Obtaining land use characteristics and spatial structures of cities have been conducted through mobile phone data. Pei et al. (2014) developed a method to cluster land usages into 5 types based on normalized hourly call volume and the total call volume (Figure 9). Dong et al. (2015) developed a method to cluster cells into traffic zones from incremental flows in each cell, and investigating a traffic zone attribute index to show land use. The findings of each study matched quite well with typical land use in the areas analysed. Louail et al. (2014) distinguished the structure of cities in Spain as monocentric and polycentric by detecting hot-spots or points of interest from the density of calls.

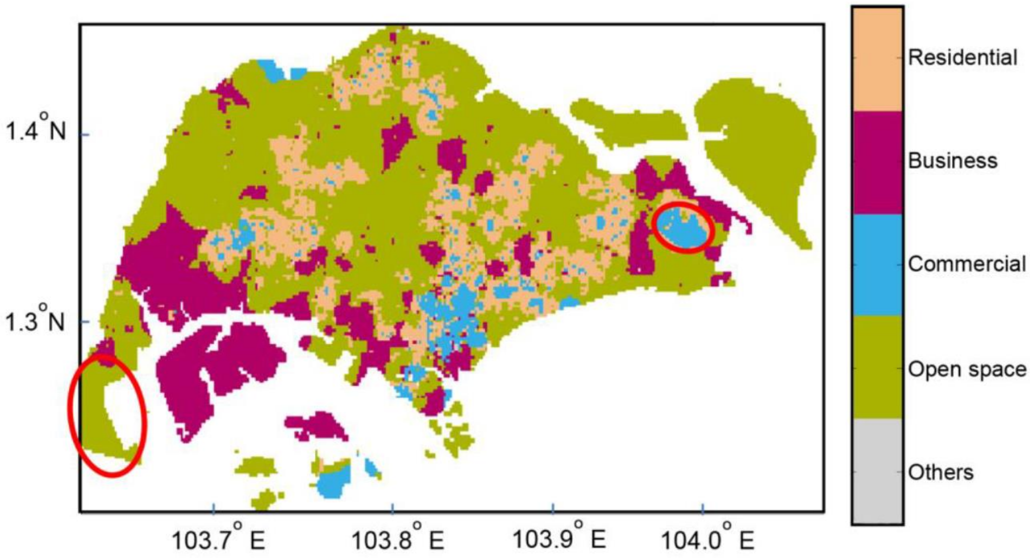


Figure 9: Clustering result for land use types in Singapore (circled area being the misclassified area) (Pei et al., 2014)

3.4 APPLICATION ON TRANSPORTATION

Human mobility is defined as the movement of individuals from one place to another. Mobile phone data can be used to study human mobility by analysing traffic flow, travel time/speed, or travel demand.

Estimation of traffic counts has been performed by Thiessenhusen et al. (2003), and Caceres et al. (2007), using handover data, and Location Update data, respectively. They described that by analysing the network topology of cells and location updates, it should be possible to consider a virtual traffic counter on the borders of location areas. However, handover is only collected from the phones that are moving between cells during a call. The limitation of Location Update data is that it is less frequent, so the accuracy and time-dependency are limited. Bar-Gera (2007) conducted experiments to measure travel time and speed using the record of handover during a phone conversation. However, it requires a scaling factor to convert mobile phone counts to vehicle counts, which is not easy on minor roads and difficult for time-dependent estimation.

Similarly, traffic congestion and density can also be estimated from sightings data, including handover data. Ratti et al. (2006) estimated the vehicle density by analysing the volume of calls from the cells that contain main roads instead of areas where people are likely to stay. The idea of monitoring the traffic using sightings data have been utilized by some public authorities (AP, 2005).

According to Caceres et al. (2008), although CDR data is limited in its ability to accurately estimate the traffic data, it widely contributes to low-cost transportation management and can be a promising alternative to the existing methods.

Table 6: List of Studies Mentioned in Section 3.4

Publication	Type of data	Contribution
Thiessenhusen et al. (2003)	Handover data	Utilizing handover data for counting traffic.
Caceres et al. (2007)	Location Update	Utilizing location update data for counting traffic through simulation.
Bar-Gera (2007)	Handover data	Measuring travel time and speed at specific streets using handover data.
Ratti et al. (2006)	CDR	Estimating the traffic density and congestion from CDR.
Caceres et al. (2008)	CDR	Reviewing the works on traffic applications from mobile phone data.

3.5 APPLICATION ON TRAVEL DEMAND

The fact that mobile phone data enables the extraction of travel patterns from a large number of individuals has made travel demand estimation a major topic of research using mobile phone records.

Akin et al. (2002) first established a method in a simulation environment to detect home/work locations by accumulating the location and define the most stayed cell at night as home and that during daytime as workplace. This basic concept to define home/work locations for OD matrices has been used in many of the later studies. Caceres (2007) developed and conducted a simulation using the information of a LA update register. Although this method can only be applied to construct OD matrices with relatively large zones, unlike CDR, the LAU always occurs when a mobile phone enters a new LA, so the penetration rate can be expected to be high. Calabrese et al. (2011a) and Frias-Martinez et al. (2012) used actual sightings data and CDR data respectively and showed that the estimation correlates well with the Census estimates.

Calabrese et al. (2011a) differ from Frias-Martinez et al. (2012) in the point that an external OD matrix was used to identify the best parameters to construct an OD Matrix. This makes it possible for the estimated OD to adapt better to different urban areas. Meanwhile, Iqbal et al. (2014) used both CDR and limited traffic counts to estimate vehicle trajectories and finally estimated OD matrix through microsimulation in MITSIMLab. Alexander et al. (2015) and Çolak et al. (2015) showed methods to estimate the OD matrix which includes more detailed trip purposes (home, work, and other) from sightings data.

Graells-Garrido et al. (2015) used the idea of a geometric approach on transportation rules for detecting a trip. Although estimating trip purpose is not available through this method, trips can be detected instantly and is stated that real-time estimation may be possible every 15 minutes. On the other hand, Jiang et al. (2017) proposed a model that is activity-based mobility pattern using CDR. They also demonstrated that users with different phone usage patterns can still have similar travel patterns.

Among many papers, the common approaches found are the following four processes: pre-processing, stay extraction, activity inference, and mode splitting. In each section, the challenges and capabilities for each process are explained.

3.5.1 Pre-processing

Pre-processing is an important part of utilizing mobile phone data due to the noise at spatiotemporal levels.

The event-triggered data sometimes is very sparse in terms of time because of low usage. The low frequent users are often filtered out, as it is hard to extract movements from such data. Due to the filtering process, the final data is biased towards heavy users of mobile phones. One common cause of noise in mobile phone data is oscillation - a phenomenon in which the phone switches connection between BSs while staying in the same position. Wang & Chen (2018) and Wu et al. (2014) worked on processing cellular-based data and triangulated sightings data respectively, managing to remove oscillations for better estimation.

3.5.2 Stay Extraction

The extraction of the stayed locations of individuals is linked to forming the trip-ends for each OD trip, thus the accuracy of the stay-extraction is important for the spatial accuracy of OD estimation.

The basic concept of extracting a stayed location is to set a threshold for the length of time that the phone was found to be in the same location. This length of time varies from 30 minutes to several hours (Calabrese, 2011a; Graells-Garrido & García, 2015).

The spatial accuracy of the detected location heavily relies on the size of cells. The size could be slightly smaller if the location is detected by triangulation. This leads to the conclusion that the minimum spatial resolution of OD estimation is more or less equivalent to the Statistical Area Level 2 (SA2) level in Australia (Tobias Hemmings, 2016).

3.5.3 Activity Inference

The basic concept of inferring activity from mobile phone data is to extract meaningful locations of each user and making it possible to estimate the types of trips. In most cases, we could assume the location of each mobile phone user as a home if the detection is between 7 PM and 8 AM on weekdays and for the entire 24 hours during the weekends. The second most visited location during the rest of the time is identified as working place. This makes it possible to estimate the following three types of trips: Home-Work trips, Home-Other trips, and Other trips. Tobias Hemmings (2016) has concluded that a minimum of 10 days of accumulation would make these estimations reliable.

3.5.4 Mode Splitting

Mode-specific OD Matrices are usually created by conducting mode splitting of the OD Matrix. Qu et al. (2015), worked on mode splitting OD demand derived from CDR, using the traditional mode split method. Huang et al. (2019) has concluded that although this showed a good correlation with most census data, some deviations could also be predicted, which needed more investigation in future studies.

3.6 SUMMARY

Here in this chapter, the applications of mobile phone data, especially CDR data in the fields of geographical analysis and transportations have been listed and reviewed. The geographical analysis includes geographical partitioning, population distribution, identification of meaningful locations, and estimation of land use. Each field had the potential to be applied in practice. The analysis on transportation other than travel demand includes traffic count, travel speed, and traffic density. There have been many contributions of such kind using CSD, and some are being used in practice. Finally, the contributions on travel demand show that extracting travel demand from mobile phone data has become possible from both network-driven data and event-driven data. Although the trip purposes extracted from mobile phone data correlates well with the survey data, there remain challenges for estimating mode-specific OD Matrices.

Table 7: List of papers mentioned in Section 3.5

Publication	Type of data	Contribution
Akin and Sisiopiku (2002)	Periodic Update (Simulated)	Extracting trips from mobile phone data through simulation.
Caceres et al. (2007)	LA data	Extracting trips from location update data through simulation.
Calabrese (2011a)	CDR (triangulated data)	Finding correlations with the census and the extracted large-scale CDR data.
Frias-Martinez et al. (2012)	CDR (triangulated data)	Finding correlations with the census and the extracted large-scale CDR data.
Alexander et al. (2015)	CDR (triangulated data)	Estimation of OD matrices from CDR
Çolak et al. (2015)	CDR (triangulated data)	Estimation of OD matrices from CDR
Graells-Garrido and García (2015)	CDR	Extracting trips with a "geometric approach"
Iqbal et al. (2014)	CDR & traffic counts	Inferring traffic counts with OD Matrices from CDR
Jiang et al. (2017)	CDR	Extracting activities from CDR
Wang and Chen (2018)	CDR	Removing oscillations for stay extractions.
Wu et al. (2014)	Sighting Data (triangulated data)	Removing oscillations of triangulated data for stay extractions.
Qu et al. (2015)	CDR	Mode Split for intra-urban OD Matrices

Chapter 4: Synthetic CDR and Experiments using Real Data

4.1 BACKGROUND

Few studies discussed the findings from the analysis of real mobile phone data and very limited have emphasised the need for a comprehensive framework to generate synthetic CDR.

4.2 METHODOLOGY

This section presents a methodology for constructing the synthetic CDR. This requires three sources of data:

- Sample connection event data,
- Individual travel diary, and
- Cell tower locations.

Sample connection event data is needed for identifying the patterns necessary to generate the timestamps for synthetic CDR. As mentioned in Chapter 2, connection events happen when the phone is involved in a call, SMS, or an Internet connection, and being recorded on the CDR. These events can be extracted from small numbers of actual CDR, as they would be an example of how connection events happen.

Secondly, the individual travel diary that represents individual journeys needs to be collected, and this can be from any source if it can represent the journey of individuals (Household Travel Survey, collected GPS trajectories, etc.). The individual trajectory will be estimated from the individual travel diaries.

Finally, the cell phone tower locations will be needed to identify the connected cell phone tower at each timestamp.

As shown in Figure 10, this process consists of two steps. Firstly, the trajectories will be overlayed with the sample connection event data, and the list of connection events linked with its locations will be created (Figure 11). It needs to be noted that this overlapping concept stands on the assumption that the occurrence of connection events does not correlate with the movement pattern of each user. Then as the second step, the cell phone tower that was connected at each event will be estimated using the cell phone tower locations data, and the list of events will be linked with the Cell IDs, which is the estimated CDR dataset.

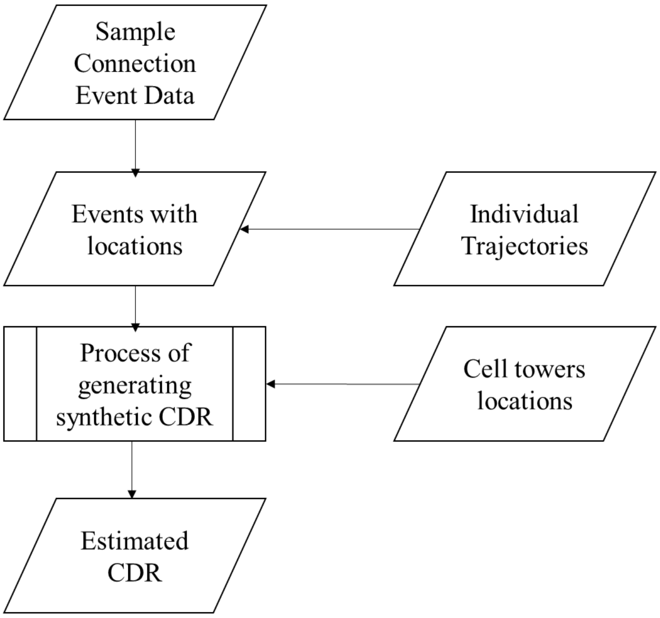


Figure 10: Generation framework of the synthetic CDR

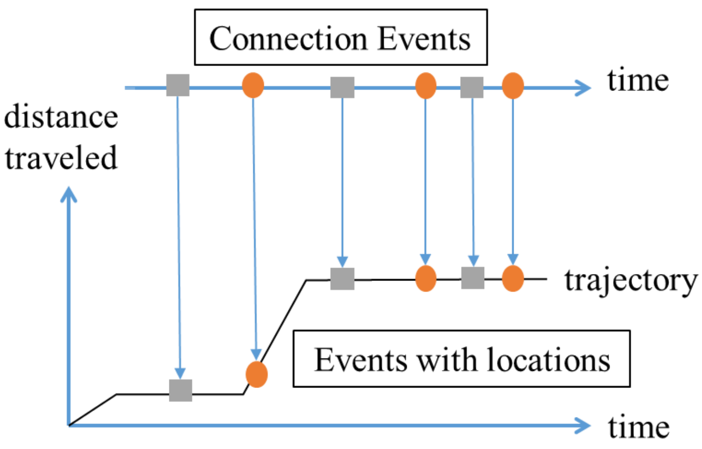


Figure 11: Description of the concept of overlaying the connection events with trajectory

Here in this experiment, 5 students from the Queensland University of Technology (QUT) collected their trajectories using the timeline function on Google Maps application on their smartphones. It was assumed that the trajectory was collected mainly with GPS. As a sample connection event data, 5 unique sample event distributions were used, all of which being the mobile phone usage, a single person disclosed, from Vodafone Australia. The cell tower locations are obtained from the Radio Frequency National Site Archive (RFNSA) website (https://www.rfnsa.com.au/), where the exact locations of all cell phone antennas are listed, assuming that when a timestamp is recorded on the CDR, the connected cell tower is the closest one from the user's estimated location. Figure 12 shows one example of data collected by the user E on one day of data that was used for the experiment.

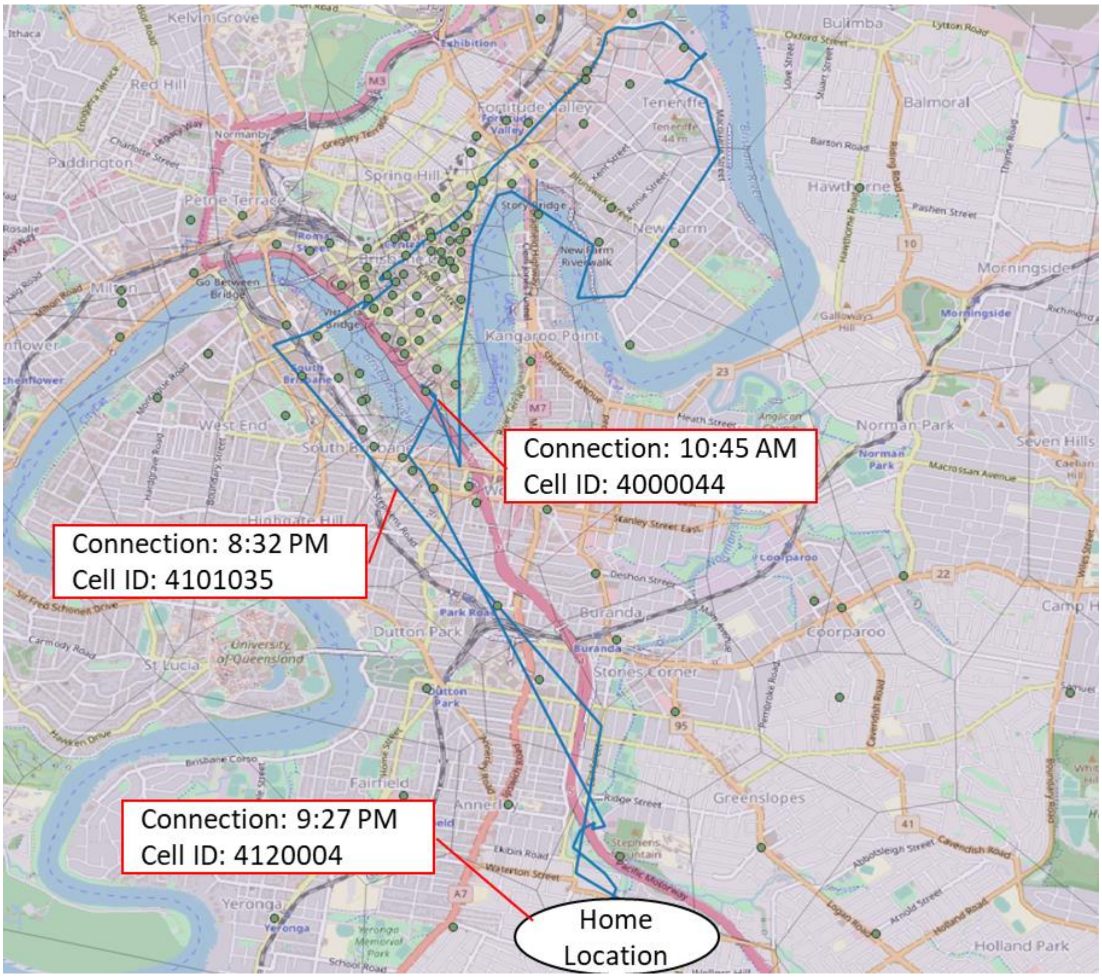


Figure 12: Trajectory of one day collected by participant E, cell tower locations of Vodafone Australia provided from RFNSA Website and estimated locations of connection events

Here in this data, the trajectory of the day, Vodafone Australia's cell tower locations, and the estimated locations of the three connection events are shown. In Figure 12, a trajectory shows where the person has moved along throughout this day, and it also shows where the three connections happened that day. For each connection, the closest tower was assumed to be the connected base station, and the ID unique to each cell is shown in the figure. Table 8shows the estimated CDR of this day. Here in this table, the phone number of the user (encrypted), type of connection (whether it's a phone, SMS, or data use), the number called (amount of data if it's a connection by data use), details of the usage, date, time, location (cell ID), duration of the call (amount of data for data use connection), and finally the charged fee of each connection was recorded. Table 9 describes the overall details of the data collected by participants A-E. The frequencies of the time stamps for each user do not depend only on each user's usage but also depends on which operators they are subscribed with, as it can be considered that each operator collects billing data differently.

Table 8: Part of Synthetic CDR relevant to trajectory shown in Figure 12

Calling Number	Usage Type	Number Called	Call Type/Description	Date	Start Time	Your Location	Duration	You Pay
61000000000	Data	Data 30715KB	National Data	12/11/2018	10:45:00 AM	4000044	30715KB	0
61000000000	Data	Data 51211KB	National Data	12/11/2018	8:32:25 PM	4101035	51212KB	0
61000000000	Data	Data 5605KB	National Data	12/11/2018	9:27:32 PM	4120004	5605KB	0

Table 9: Details of each Estimated CDRs

Participant ID	A	B	C	D	E
Number of days collected	2	6	6	7	20
Number of total timestamps	23	424	74	118	137
Network Operator	Vodafone AU	Optus	Optus	Optus	Vodafone AU

4.3 EXPERIMENT

4.3.1 Methodology

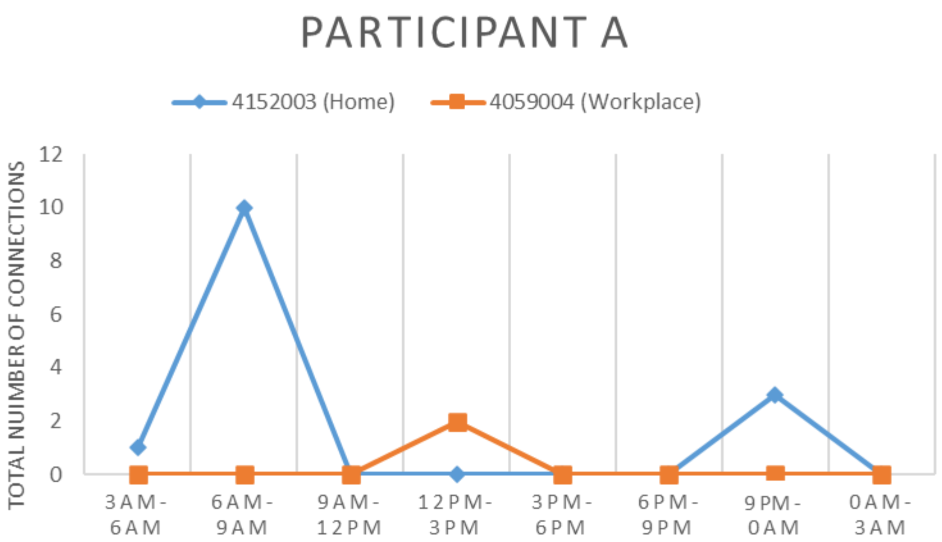
Here in this experiment, the accuracy of extracting individuals' meaningful locations was tested. The meaningful locations refer to where the individuals stay occasionally such as home, work, or school locations. This experiment was conducted in two steps. As the first step, the cells which had frequent detections by an individual were listed up as they are likely candidates of meaningful locations. And as the second step, the time of day of the detections in each cell was analysed, so that they could be identified between home, work/school, or other locations.

4.3.2 Results

Table 10 is the list of candidate cells of individual's meaningful locations. The cells that were selected are the ones that had connection events most frequently (more than once), and the number of connections with each cell is also shown. The cell IDs are what is shown on the RFNSA website. The number of connections in each of these cells regarding the time of day is shown in Figure 13.

Table 10: List of Individuals' Meaningful Locations

Participant	A	B	C	D	E
Cell No	Cell ID (Connections)	Cell ID (Connections)	Cell ID (Connections)	Cell ID (Connections)	Cell ID (Connections)
1st cell	4152003 (12)	4115005 (292)	4000123 (9)	4101032 (93)	4120004 (54)
2nd cell	4059004 (2)	4000108 (78)	4000004 (7)	4101053 (11)	4000044 (40)
3rd cell	-	4115003 (54)	4000044 (6)	4074006 (6)	4107004 (9)
4th cell	-	-	4000120 (5)	4000121 (3)	4064001 (7)



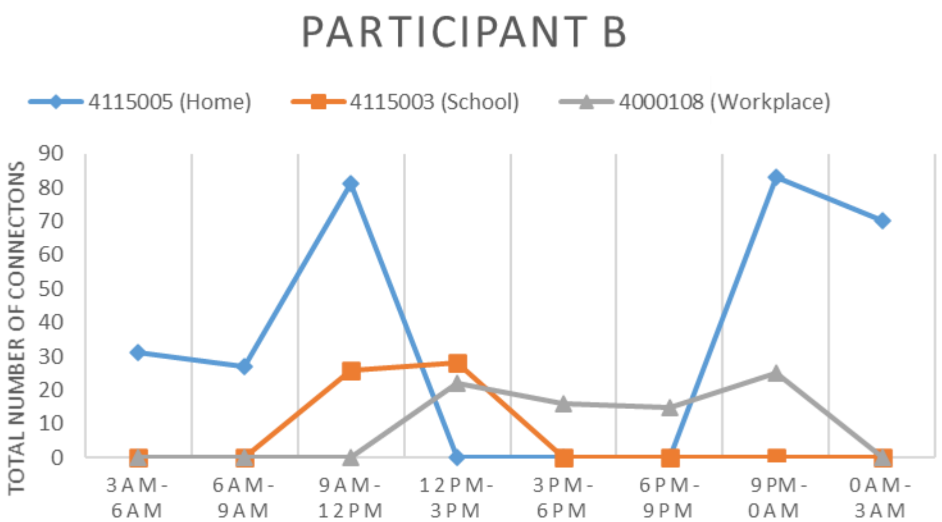


Figure 13: The number of connections in each of the candidate cells shown in Table 10

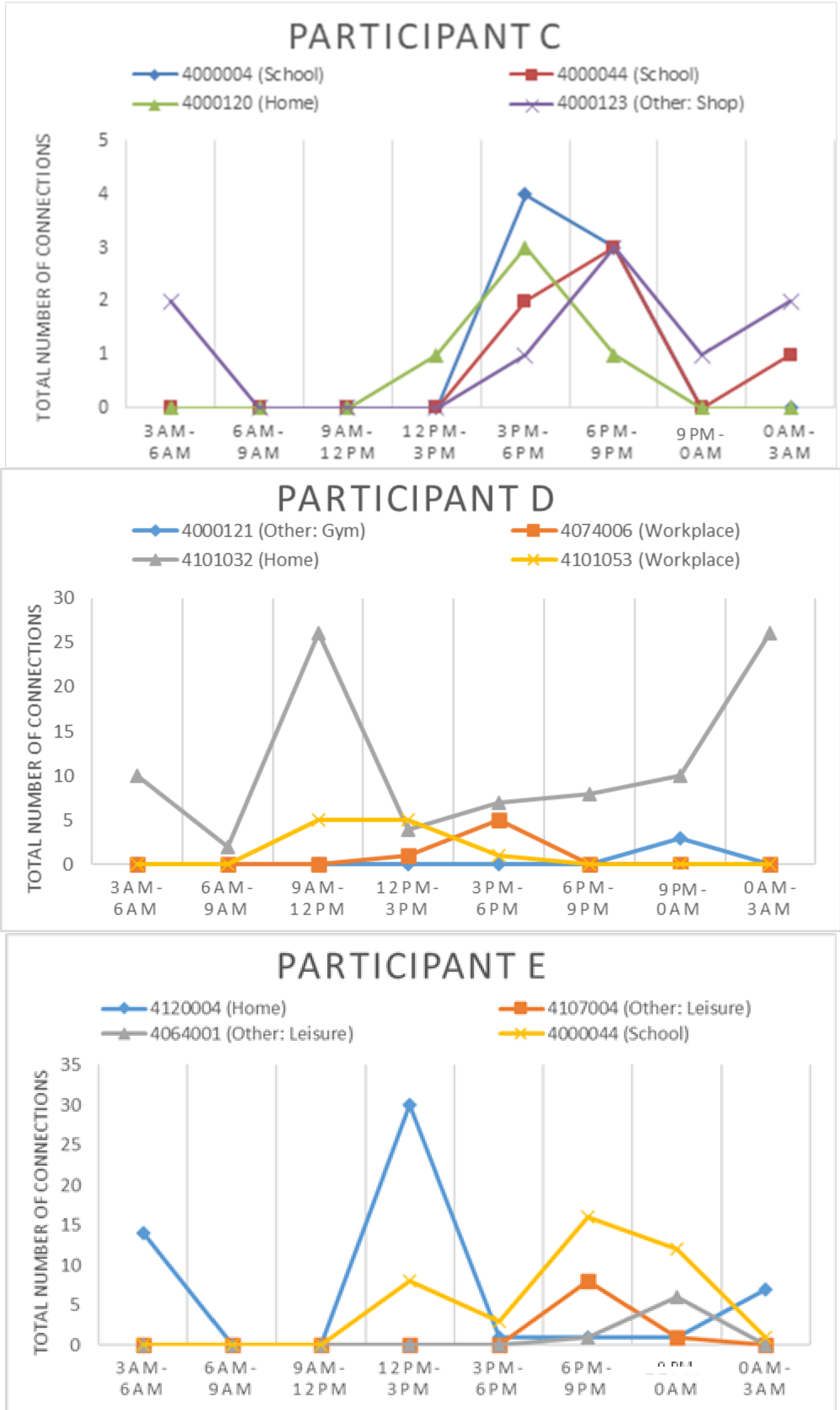


Figure 13: The number of connections in each of the candidate cells shown in Table 10 (continued)

4.4 DISCUSSION FROM THE RESULTS

From the results of the experiment, it has been shown that the tendency of connections near home location being at night and those in work/school locations being daytime can be observed. Thus, estimation of individuals' meaningful location is possible to some extent, in terms of extracting its home, work, and other locations. This high accuracy of estimating meaningful locations of individuals is one significance of CDR compared to other sources of data, as the data tracks the individual for several days and makes it possible to extract the travel patterns.

4.5 APPLICATION OF THE SYNTHETIC CDR

For synthetic CDRs to be usable for further applications, overcoming the issue of inaccessibility and incomparability, it requires to be rich in its amount of data and be less biased. This means that a method to generate a large number of unique and less-biased synthetic CDRs is required. This may become possible through techniques such as Monte Carlo simulation and creating a larger number of less biased and unique event distributions from small sample event distributions. Once this rich synthetic CDR is possible to be made, it is expected that it can be used to simulate the generation process of CDR in an agent-based simulation platform. This would help establish a synthetic "benchmark data" that can be used for various CDR applications, as well as giving opportunities for people without full access to it.

However, three points that remain as issues. Firstly, this process of overlaying the trajectories with event distributions relies on the assumption that the occurrence of trips and connection events are independent of each other, and this may cause a bias. Secondly, we assume that mobile phones are connected to the antennas that have the shortest Euclidian distance, which is not always the case (Trevisani & Vitaletti, 2004). So, a better approach needs to be developed to identify the connected towers. Finally, individuals use their phones differently, and heavy and light users may also need to be considered in the event distributions because the temporal sparsity affects the estimation accuracy of the locations (Schlaich et al., 2010). Thus, verification of the synthetic CDR in terms of these three points still needs to be explored.

Chapter 5: Conclusion

Researchers have constantly been analysing mobile phone data and making new applications for better use. Thus, reviewing the state-of-the-art, revealing the remaining challenges and limitations, and finding a pathway for further applications should contribute to the future analysis of this data. This is the main motivation of this research and the findings from each chapter are summarised below.

In Chapter 2, various mobile phone data have been reviewed for their fundamental understanding. Mobile phone data includes data for billing purposes (CDR and IPDR) and management purposes (Triangulated records, LAU Register data, and Handover data). It can be concluded that the trips can be extracted regardless of the type of data, but bias can remain.

In Chapter 3, the applications of mobile phone data on mobility studies and related fields have been reviewed for revealing the remaining challenges. The fields of applications included transportation and geographical analysis as a related field. Each field included in the category of geographical analysis had the potential to be applied in practice. The data that is extracted for management purposes are often used for traffic state estimation, which can be associated with the estimation of travel speed and travel time estimation, and this has been used in practice. Regarding the travel demand, there could be seen that there is a high performance with extracting OD matrices with trip purposes, but extracting modes are still being challenged.

In this Chapter 4, a framework for synthetic CDR data by simulating the generation process of CDR data has also been proposed. The construction of synthetic CDR data aims to make this data a “benchmark data” so that various contributions can be compared to each other, as well as giving opportunities to have access to the CDR data, as the actual CDR data is often expensive. The results of the simple experiment were supporting the usability of the data.

The significance of this research is that it brings an understanding of the mobile phone data and its applications, as well as its challenges on the applications and its solution. The mobile phone data, which can be passively extracted from the mobile phone network, such as Call Detail Records (CDR), are rich in the amount of data containing human behaviour. This has enabled the researchers to extract various information related to human behaviour, including travel demand. However, extracting the mode-specific trips from CDR remains a challenge. Also, the bias that the mobile phone data can contain remains to be explored.

This research is focused only the applications of mobile phone data for transportation application including geographical analysis. Applications of the other fields are not considered in this study. The mobile phone data is defined as the set of data that is passively generated from the mobile phone network. This study does not include other types of data that are collected intensively, such as GPS data or additional phone apps used for data collection.

Some of the findings from this research were presented in the following paper "Insights on the mobile phone data applications" in the World Conference on Transport Research - WCTR 2019, Mumbai, India (Suzudo et al., 2019). As future research, the synthetic CDR can be further explored, so that various types of traffic analysis including the extraction of mode-specific trips and their bias can be simulated.

Bibliography

ACMA. (2019). *Communications Report 2018-19*.

Adam, A., Delvenne, J.-C., & Thomas, I. (2017). Cartography of interaction fields in and around Brussels: commuting, moves and telephone calls. *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles/Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel/The e-journal for academic research on Brussels*.

Advani, C., Bhaskar, A., Haque, M. M., & Cholette, M. E. (2021). STATER: Slit-Based Trajectory Reconstruction for Dense Urban Network With Overlapping Bluetooth Scanning Zones. *IEEE Transactions on Intelligent Transportation Systems*.

Ahas, R., Silm, S., Järv, O., Saluveer, E., & Tiru, M. (2010). Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of urban technology, 17*(1), 3-27.

Akin, D., & Sisiopiku, V. (2002). Estimating Origin-Destination Matrices Using Location Information from Cellular Phones. *Proc. NARSC RSAI, Puerto Rico, USA.*

Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, *58*, 240-250.

Anda, C., Erath, A., & Fourie, P. J. (2017). Transport modelling in the age of big data. *International Journal of Urban Sciences*, *21*(sup1), 19-42.

AP. (2005). Tracking cell phones for real time traffic data. *Associated Press*. http://www.nbcnews.com/id/9698139/ns/technology_and_science-wireless/t/tracking-cell-phones-real-time-traffic-data/

Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies, 15*(6), 380-391.

Barcelö, J., Montero, L., Marqués, L., & Carmona, C. (2010). Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation research record, 2175*(1), 19-27.

Behara, K., Bhaskar, A., & Chung, E. (2018). Classification of typical Bluetooth OD matrices based on structural similarity of travel patterns: Case study on Brisbane city. Proceedings of the Annual Meeting The Transportation Research Board (TRB) 97th Annual Meeting, Behara, K. N., Bhaskar, A., & Chung, E. (2020a). Geographical window based structural similarity index for origin-destination matrices comparison. *Journal of Intelligent Transportation Systems*, 1-22.

Behara, K. N., Bhaskar, A., & Chung, E. (2020b). A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*, *111*, 513-530.

Behara, K. N., Bhaskar, A., & Chung, E. (2020c). A novel methodology to assimilate sub-path flows in bi-level OD matrix estimation process. *IEEE Transactions on Intelligent Transportation Systems*, 1-11. https://doi.org/doi: 10.1109/TITS.2020.2998475

Behara, K. N., Bhaskar, A., & Chung, E. (2021). Single-level approach to estimate origin-destination matrix: exploiting turning proportions and partial OD flows. *Transportation Letters*, 1-12.

Bhaskar, A., & Chung, E. (2013). Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. *Transportation Research Part C: Emerging Technologies*, *37*, 42-72.

BITRE. (2014). *New traffic data sources – An overview*. https://www.bitre.gov.au/sites/default/files/2019-12/NewDataSources-BackgroundPaper-April%202014.pdf

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ data science*, *4*(1), 10.

Caceres, N., Wideberg, J., & Benitez, F. (2007). Deriving origin–destination data from a mobile phone network. *IET Intelligent Transport Systems*, *1*(1), 15-26.

Caceres, N., Wideberg, J., & Benitez, F. G. (2008). Review of traffic data estimations extracted from cellular networks. *IET Intelligent Transport Systems*, *2*(3), 179-192.

Calabrese, F. (2011a). *Urban sensing using mobile phone network data* https://researcher.watson.ibm.com/researcher/files/ie-FCALABRE/Urban%20sensing%20using%20mobile%20phone%20network%20data.pdf

Calabrese, F. (2011b). *Urban sensing using mobile phone network data Ubicomp 2011 Tutorial* Retrieved 27/07 from https://researcher.watson.ibm.com/researcher/files/ie-FCALABRE/Urban%20sensing%20using%20mobile%20phone%20network%20data.pdf

Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, *10*(4), 36-44.

Chandrasekar, P. (2015). Big data and transport modelling: opportunities and challeges. *Int J Appl Eng Res, 10*(17), 38038-38044.

Chen, C., Bian, L., & Ma, J. (2014). From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies, 46*, 326-337.

Chi, G., Thill, J. C., Tong, D., Shi, L., & Liu, Y. (2016). Uncovering regional characteristics from mobile phone data: A network science approach. *Papers in Regional Science, 95*(3), 613-631.

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., & González, M. C. (2015). Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record, 2526*(1), 126-135.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences, 111*(45), 15888-15893.

Dong, H., Wu, M., Ding, X., Chu, L., Jia, L., Qin, Y., & Zhou, X. (2015). Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies, 58*, 278-291.

Frias-Martinez, V., Soguero, C., & Frias-Martinez, E. (2012). Estimation of urban commuting patterns using cellphone network data. Proceedings of the ACM SIGKDD International Workshop on Urban Computing,

Fukumoto, J., & Okamoto, Y. (2012). Delineation of Functional Regions by Combining Community Finding Method with Spatial Interaction Model (Japanese Title: コミュニティ抽出法と空間相互作用モデルを組み合わせた機能地域区分手法の提案). *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management), 68*(5), I_427-I_436.

Ghadialy, Z. (2017). *Macrocells, Small Cells & Hetnets Tutorial*. Retrieved July 26th from https://blog.3g4g.co.uk/2017/12/macrocells-small-cells-hetnets-tutorial.html

Graells-Garrido, E., & García, J. (2015). Visual exploration of urban dynamics using mobile data. International Conference on Ubiquitous Computing and Ambient Intelligence,

Houlton, S. (2011). *German politician reveals six months of private mobile phone data*. Retrieved 18/08 from https://www.dw.com/en/german-politician-reveals-six-months-of-private-mobile-phone-data/a-14960858

Huang, H., Cheng, Y., & Weibel, R. (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies, 101*, 297-312.

Hussain, E., Behara, K. N., Bhaskar, A., & Chung, E. (2021). A Framework for the Comparative Analysis of Multi-Modal Travel Demand: Case Study on Brisbane Network. *IEEE Transactions on Intelligent Transportation Systems*.

Hussain, E., Bhaskar, A., & Chung, E. (2021a). A novel origin destination based transit supply index: Exploiting the opportunities with big transit data. *Journal of Transport Geography, 93*, 103040.

Hussain, E., Bhaskar, A., & Chung, E. (2021b). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies, 125*, 103044.

Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies, 40*, 63-74.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. International Conference on Pervasive Computing,

ITU. (2017). *Call Detail Record (CDR) Analysis: Sierra Leone*.

ITU. (2019). *10 Network Coverage*. International Telecommunication Union https://itu.foleon.com/itu/measuring-digital-development/network-coverage/

Jaume, B., & Montero, L. (2015). An integrated Computational framework for the estimation of dynamic OD trip matrices. 2015 IEEE 18th International Conference on Intelligent Transportation Systems,

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

Kieu, L.-M., Bhaskar, A., & Chung, E. (2015). A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies, 58*, 193-207.

Louail, T., Lenormand, M., Ros, O. G. C., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific reports, 4*, 5276.

Michau, G., Pustelnik, N., Borgnat, P., Abry, P., Nantes, A., Bhaskar, A., & Chung, E. (2017). A primal-dual algorithm for link dependent origin destination matrix estimation. IEEE Transactions on Signal and Information Processing over Networks,

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.-L., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science, 28*(9), 1988-2007.

Qu, Y., Gong, H., & Wang, P. (2015). Transportation mode split with mobile phone data. 2015 IEEE 18th international conference on intelligent transportation systems,

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design, 33*(5), 727-748. https://doi.org/10.1068/b32047

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., & Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS one, 5*(12), e14248.

Rojas IV, M. B., Sadeghvaziri, E., & Jin, X. (2016). Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. *Transportation Research Record, 2563*(1), 71-79.

Schlaich, J., Otterstätter, T., & Friedrich, M. (2010). Generating trajectories from mobile phone data. Proceedings of the 89th annual meeting compendium of papers, transportation research board of the national academies,

Simini, F., Barlacchi, G., Pellungrini, R., & Pappalardo, L. (2019). Human Mobility from theory to practice: Data, Models and Applications.

Steenbruggen, J., Borzacchiello, M. T., Nijkamp, P., & Scholten, H. (2013). Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. *GeoJournal, 78*(2), 223-243.

Suzudo, M., Behara, K. N. S., & Bhaskar, A. (2019). Insights on the mobile phone data applications. World Conference on Transport Research (WCTR), Mumbai, India.

Telebeansolutions. (2015). *GSM Location area update*. Retrieved 27/07 from https://www.slideshare.net/Telebeans/location-area-update#:~:text=2.,where%20the%20MS%20is%20located.&text=This%20is%20how%20the%20MS,Identity%20of%20the%20serving%20cell.

Thiessenhusen, K., Schafer, R., & Lang, T. (2003). Traffic data from cell phones: a comparison with loops and probe vehicle data. *Institute of Transport Research, German Aerospace Center, Germany*.

Tobias Hemmings, C. G. (2016). *Utilising Mobile Network Data for Transport Modelling*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662948/utilising-mobile-network-data.pdf

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources [Article]. *Transportation Research Part C: Emerging Technologies, 58*, 162-177. https://doi.org/10.1016/j.trc.2015.04.022

Trevisani, E., & Vitaletti, A. (2004). Cell-ID location technique, limits and benefits: an experimental study. Sixth IEEE workshop on mobile computing systems and applications,

von Mörner, M. (2017). Application of call detail records-chances and obstacles. *Transportation research procedia, 25*, 2233-2241.

Wang, F., & Chen, C. (2018). On data processing required to derive mobility patterns from passively-generated mobile phone data. *Transportation Research Part C: Emerging Technologies, 87*, 58-74.

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society, 11*, 141-155.

Wu, W., Wang, Y., Gomes, J. B., Anh, D. T., Antonatos, S., Xue, M., Yang, P., Yap, G. E., Li, X., & Krishnaswamy, S. (2014). Oscillation resolution for mobile phone cellular tower data to enable mobility modelling. 2014 IEEE 15th International Conference on Mobile Data Management,