Identification of daily human activity patterns from mobile CDR in Yangon City

Win War HLAING[a], Kyaing[b], Rubel DAS[c] Shigeru NAKAMURA[d]

[a,b]*Yangon Technological University, Yangon, Myanmar;*
[a]*E-mail: winwarhlaing7@gmail.com*
[b]*E-mail: kyaingkyaing63@gmail.com*
[cd]*Research and Development Center, Nippon Koei Co. Ltd., Tsukuba, Japan;*
[c]*E-mail: das-rb@n-koei.jp*
[d]*Email: nakamura-sh@n-koei.jp*

Abstract: Understanding the patterns of daily human activities is essential for activity-based travel demand management and future transport planning. The ubiquitous call detail records (CDR) provide an opportunity to explore individual users' itineraries. This paper aims to identify the human activity pattern and trip purposes of each user based on their cell phone usage. Trips were computed by observing consecutive stay points. Then, trip purposes were established by analyzing the types of activities at the destination locations. This paper presents a relationship between meaningful stay points and points of interest (POI) such as home, work, shopping, and recreation. We compared the results of trip purposes created from CDR data with those form a ground truth survey of users' trips. The methodology was demonstrated using 2.3 billion call records of 2.2 million unique mobile users. The results show that the CDR data can be reliably used for trip purpose classification.

Keywords: CDR, POI, Trip Purposes, Activity Pattern

1. INTRODUCTION

Characterizing trip activities, such as duration, purpose, and time of day, is vital for future urban and transport planning, public facility management, and travel demand management. In addition, evaluation of the effects of trips relies on understanding how individuals travel and conduct activities and when they perform these activities. These issues have been the major focus areas of city and transportation planners, who need to solve the challenges of changing trip patterns by adjusting transportation network capacity through creation of new infrastructure or by managing travel demand. As the formation of a trip might depend on the traits of travel mode and destination location, the investment in these sectors could influence the total number of trips. Therefore, the quantification of trip patterns, which is traditionally computed from travel surveys, is a key step for transportation planning models such as the four-step model and space syntax. Although large-scale travel surveys can provide detailed information to transportation planners, these surveys are infrequent, expensive, and lengthy; this limits their use for transportation planning. Data from smartphones and wearable devices could be a feasible solution to overcome these difficulties. Owing to the rapid development of mobile phone technologies and cellular network systems, these technologies are widely used. Call detail records (CDRs) contain a series of consecutive spatial locations of anonymized users; the access time and locations of cellular towers can be observed, which represents the daily space-time trajectory of a particular user. The significant advantages of using CDR data are the large

user base and rich spatiotemporal information about human mobility patterns. Mobile phone users leave footprints of their interactions with the environment and digital infrastructure (Girardin et al., 2009). With affordable computing technology, CDR provides us new opportunities to investigate, visualize, and predict urban activities.

Compared to other location tracking methods, such as GPS-equipped vehicles, location-aware mobile applications, and wearable devices, the application of CDR for transportation planning could be vast. The tracking methods other than CDR have limitations in terms of usage in specific areas, purposes, and user groups. As telecommunication service providers maintain data for billing purposes (Dash et al., 2015), CDR becomes the largest subset of mobile network big data, and these data are readily available. Therefore, mobile CDR data provide a great opportunity to better understand dynamic human mobility patterns.

Yangon, the major commercial city in Myanmar, was selected for the case study in this research project. This is because the population and use of vehicles are rapidly growing, leading to traffic anomalies such as congestion and accidents. In addition, the demand for public transportation in this city does not match its supply. To improve the performance of the transportation system in this city, the analysis of trip patterns is essential.

This study concentrates on extracting origin-destination (OD) trip patterns by purpose from CDR in the Yangon urban area. The contributions of this paper are twofold. First, integration of land-use data and CDR data to classify trip purposes is a novel approach. Trip purposes were defined by comparing the activity location of participants to the land-use database. A probabilistic model was proposed for matching the CDR data and land use. The other contribution of this study is that it used a database with 2.3 billion records and 2.2 million unique mobile users; this is the largest database used according to our knowledge. The results of our proposed model were validated with true field data.

The remainder of this paper is organized as follows. In Section 2, we present the background related to this study. Section 3 describes the reason for selection of Yangon for the case study. Section 4 briefly describes the data used in this study and the method of extracting information on the meaningful points, stay point extraction, and origin-destination pairs with the trip by purpose and time of day. Section 5 covers the results and discussion of the trip by purpose with the time of day, and trip distribution with activity patterns. Finally, Section 6 covers conclusion of this study by summarizing the findings.

2. LITERATURE REVIEW

For decades, considerable effort has been devoted to identifying and characterizing the dynamics of human travel patterns, which are measured by various aspects such as daily trip frequency, trip purpose, departure time, travel duration, travel distance, travel mode, trip sequence, and trip destination. A wide range of influencing factors have been examined to understand the travel behavior. These factors are demographic and socio-economic attributes (Boarnet and Hsu, 2015), built environment (He, 2011), and personal preferences (Van Acker et al., 2016). These analyses are largely based on data from traditional travel surveys, which have evolved from questionnaires to recalled travel diaries. However, these types of surveys have shortcomings, such as high cost of collection, small sampling rates, short survey duration, under-reporting (Shen and Stopher, 2014), and coarse spatiotemporal resolution (Calabrese et al., 2013; Diao et al., 2015). Parallel to the evolution of traditional travel survey, various sources of data with unprecedented volume, termed as “big data” have emerged; among them, mobile big data are the most widely used data source. Owing to the rapid development of mobile technology, this data source has become an increasingly effective sensor for our daily

whereabouts (Lane et al., 2010). Moreover, Song et al. (2010) found that by using CDR data, it is theoretically possible to predict individual movements of users with accuracy as high as 93%.

Mobile phone data have been widely studied in various research fields such as disaster response (Lu et al., 2012), health (Wesolowski et al., 2012), and socioeconomics (Eagle et al., 2010). For a decade, the application of mobile big data has become prevalent in long-term guidance and short-term strategies for urban planning and transportation development. Many studies have discussed the application of mobile big data in activity-travel behavior and mobility patterns (Jiang, et al., 2017; Demissie et al., 2019; Batran et al., 2018; Chen et al., 2016; Kyaing et al., 2017), inferring travel demand estimation (Gundlegård et al., 2016), and trip distribution (Demissie et al., 2018). The mobile data from Korea Telecom was used to determine the movement of people at night to enhance the quality of public transport at night (ITU-T, 2013). Kung et al. (2014) analyzed home-to-work commuting patterns using CDR. Pei et al. (2014) derived land-use information from mobile phone data and characterized land-use classification. Louail et al. (2015) proposed a methodology for extracting commuting OD matrices for 31 Spanish cities from CDR data. Lee et al. (2018) proposed a model for the urban spatiotemporal analysis of South Korean cities. While previous studies have shown that mobile big data can be a fertile source for analyzing human travel patterns, they have not been able to fully address the spatiotemporal locations of population and activity patterns. In our study, we established a relationship between land use and the number of trips.

To define daily human activities from mobile big data, rule-based algorithms are commonly used. As CDR data do not contain information on the user's age or income, the success of CDR data usage depends on setting reliable rules for identifying trip purpose. For example, if a user frequently accesses a particular cellular tower during the night, the tower location will be considered as the home of the user. Similarly, if the activities are observed during working hours, they will be labeled as working in an office. Other cases can be defined as other activities in general (Isaacman et el., 2011; Colak et al., 2015). The most widely identified trip types are home-based-work trips, home-based-other trips, non-home-based-work trips, and non-home-based-other-trips (Colak et al., 2015). However, these studies failed to incorporate cases such as shopping trips, recreational trips, and treatment trips. Therefore, existing studies on the selection of activities based on mobile data have limited applications. We have expanded the existing studies by incorporating the details of trips computed from mobile big data and trip purpose computed from land-use data.

Although many studies have extracted meaningful information about human mobility, the use of CDR data in computing OD tables has been limited. One of the aims of computing people's daily activities is to establish OD tables for traffic analysis. Guan et al. (2020) used sophisticated GPS data of the mobile user, which is available only in aggregate format, to detect the origin locations of visitors to a particular park in Tokyo. This approach overcomes the limitation of CDR data in which the location of the base tower is used to identify the position of mobile users. Wang et al. (2012) used probe vehicle data to validate the road usage information inferred from CDR data. The seminal paper by Iqbal et al. (2014) used a microsimulation model to generate an OD table from CDR data. The area selected for their study was Dhaka, Bangladesh. They validated their model by collecting small ground truth data. In our study, after extracting the trip purpose from CDR data, we computed the OD table by aggregating trips that start and end at the same traffic analysis zone. Based on the work by Alexander et al. (2015), we used scaling factors to prepare the final OD table. In this study, we used Japan International Corporation Agency (JICA, 2014) survey data to validate our proposed model, which differs from the above-mentioned studies.

3. STUDY AREA

Yangon is the most important commercial center of Myanmar; therefore, it has become a prominent destination in transportation. There are 45 townships in Yangon, which are grouped by different subregions such as central business distinct (CBD), inner city, outer city, old suburbs, and new suburbs. In this study, we selected 33 townships that fall under the jurisdiction of the Yangon City Development Committee (YCDC). A population of 5.211 million resides in 33 townships in the urban areas. Yangon is one of the central transportation regions, and traffic problems occur more frequently in these regions than in the other townships. Presently, many people are moving to urban areas for better education and health care. Owing to rapid urbanization, Yangon city is facing traffic congestion, which causes a waste of time and money; therefore, the urban transport system needs to be improved. The maps of Yangon and the selected 33 townships are shown in Figure 1.

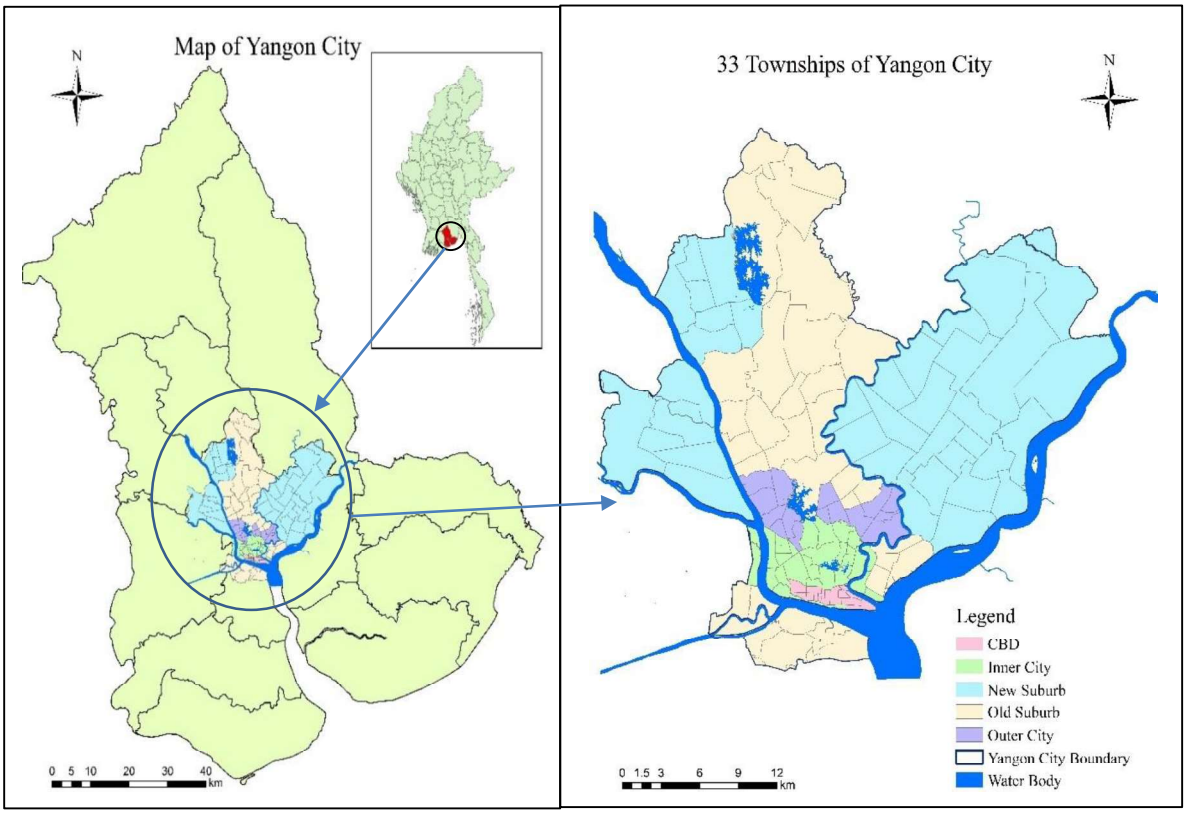


Figure 1. Location of the study area

4. METHODOLOGY

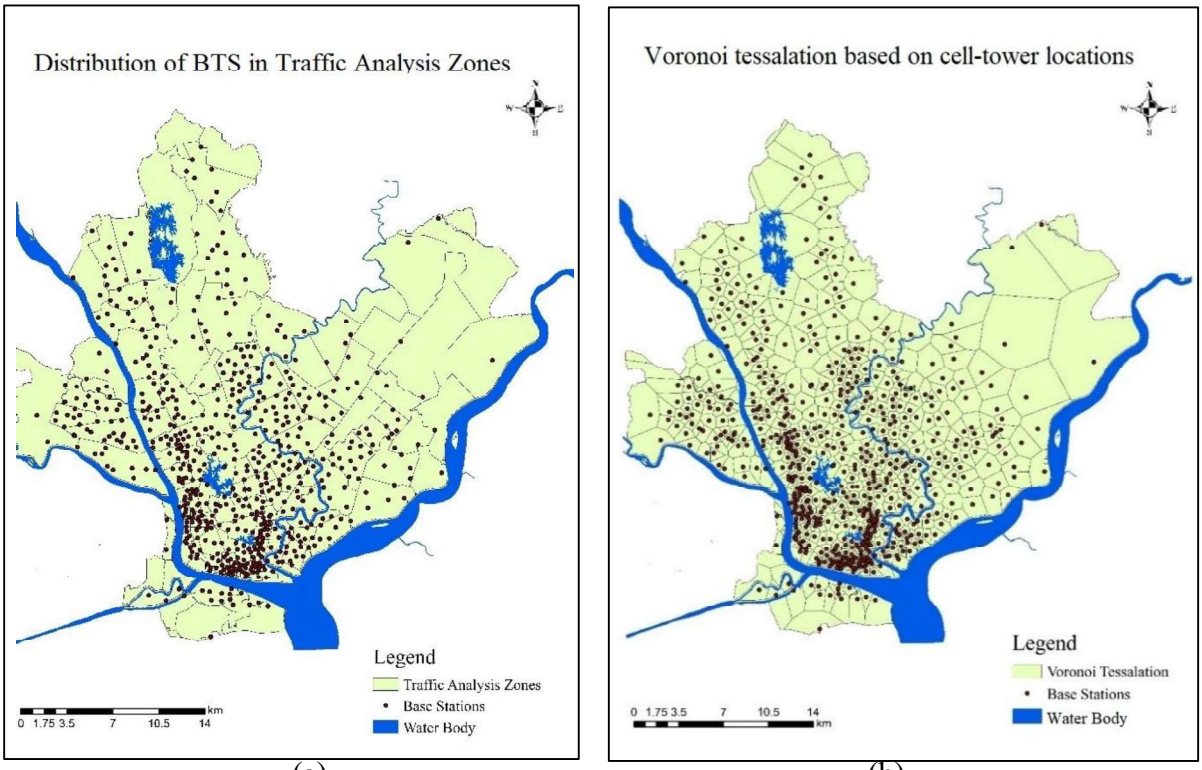
4.1 Data

CDR data were the key inputs in this study. In addition, JICA survey data and land-use data were also used. The CDR data were used to estimate origin-destination trips, and the land-use data were used to define activities at each location for the classification of trip by purpose and human activity patterns.

4.1.1 CDR Data

The CDR dataset contains more than 2.3 billion anonymized mobile phone traces made by 2.2 million unique mobile users in the Yangon urban area for one week, from December 1, 2015, to December 7, 2015. The largest telecommunication operator in Myanmar, Myanmar Post and Telecommunications (MPT), shared the CDR data with us. From the data, the locations of 657 base stations and 57 dead base stations were identified. This database has many columns such as Person ID, CELLID, event, duration, date, and time of each record. Lwin et al. (2018) also utilized similar database for computing hourly link flow computation.

The CDR data include two types of data files: data files representing internet or message usage, and voice files representing the activities with phone calls. The MPT has the highest market share (54%) in Myanmar. The BTS towers in traffic analysis zones and Voronoi tessellations based on cell-tower locations are shown in Figures 2 (a) and 2 (b), respectively.



(a) (b)

Figure 2. Distribution of BTS in (a) Traffic Analysis Zones and (b) Voronoi tessellation

4.1.2 Survey Data

To validate the estimated results from the CDR data, the 2014 Yangon Urban Transport Master Plan (YUTRA) person trip survey data were collected by the JICA. It contains 45,000 users and travel information such as trip arrival and departure times, origin and destination of trips and trip purpose. In this study, we used travel time of day, trip purpose, travel distance, and land-use information from this survey data and compared the results from the CDR data. In addition, the population from the census tracts was used to build expansion factors to transfer the information from estimated mobile phone users' travel to person-trips.

4.1.3 Points of Interest (POIs)

To establish the relationship between trips and their purposes, such as school, shopping, and social activities, detailed land-use data were incorporated into this study framework. The

facilities around the visited places were used to acquire information regarding the activities. The POI database is collected from https://www.openstreetmap.org. It provides the geographical coordinates of the POI, i.e., latitude and longitude, and the type of category labels around the location upon query. This POI information is useful in the construction of stay points to label the type of activities as "Ground truth".

4.2 Methods

In this study, human activity patterns and trips by purpose were analyzed using mobile CDR in Yangon. An overview of the research flow for this analysis is presented in Figure 3.

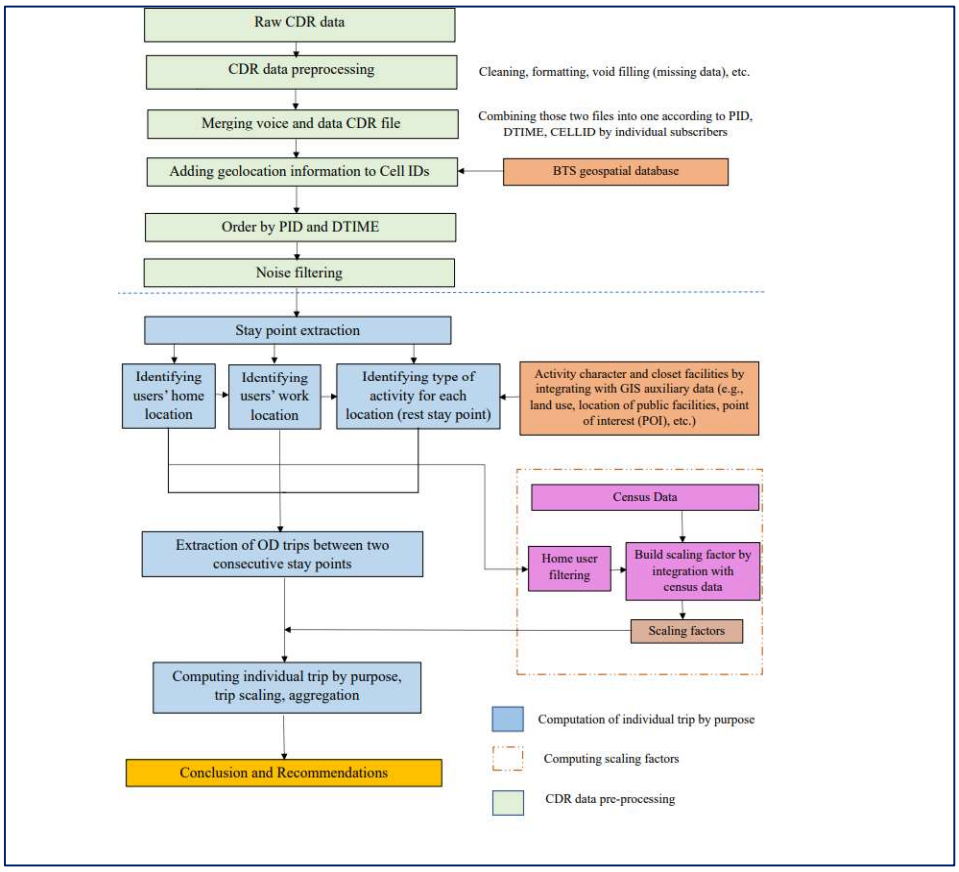


Figure 3. Overview of the proposed system

4.2.1 Preprocessing of CDR Data

Data cleaning refers to of removing unnecessary, missing, or inconsistent data. Some records may contain missing information, wrong coordinates, wrong township names, and other records whose CELLIDs may not exist in the BTS Geospatial database. Such records were removed.

Additionally, some traces always use the same CELLID and thus cannot be used for making trips. Such traces were removed.

In the second step, filtered data and voice files were merged, and the combined database was grouped using Person-ID (PID) that is a unique number of each user. Note that, the PID is an encrypted number of a user therefore specific user's identity was not addressed. Moreover,

the BTS geospatial databases, containing the latitude–longitude information of CELLID, were added to the combined file. Figure 4 shows an example of the data and voice CDR data.

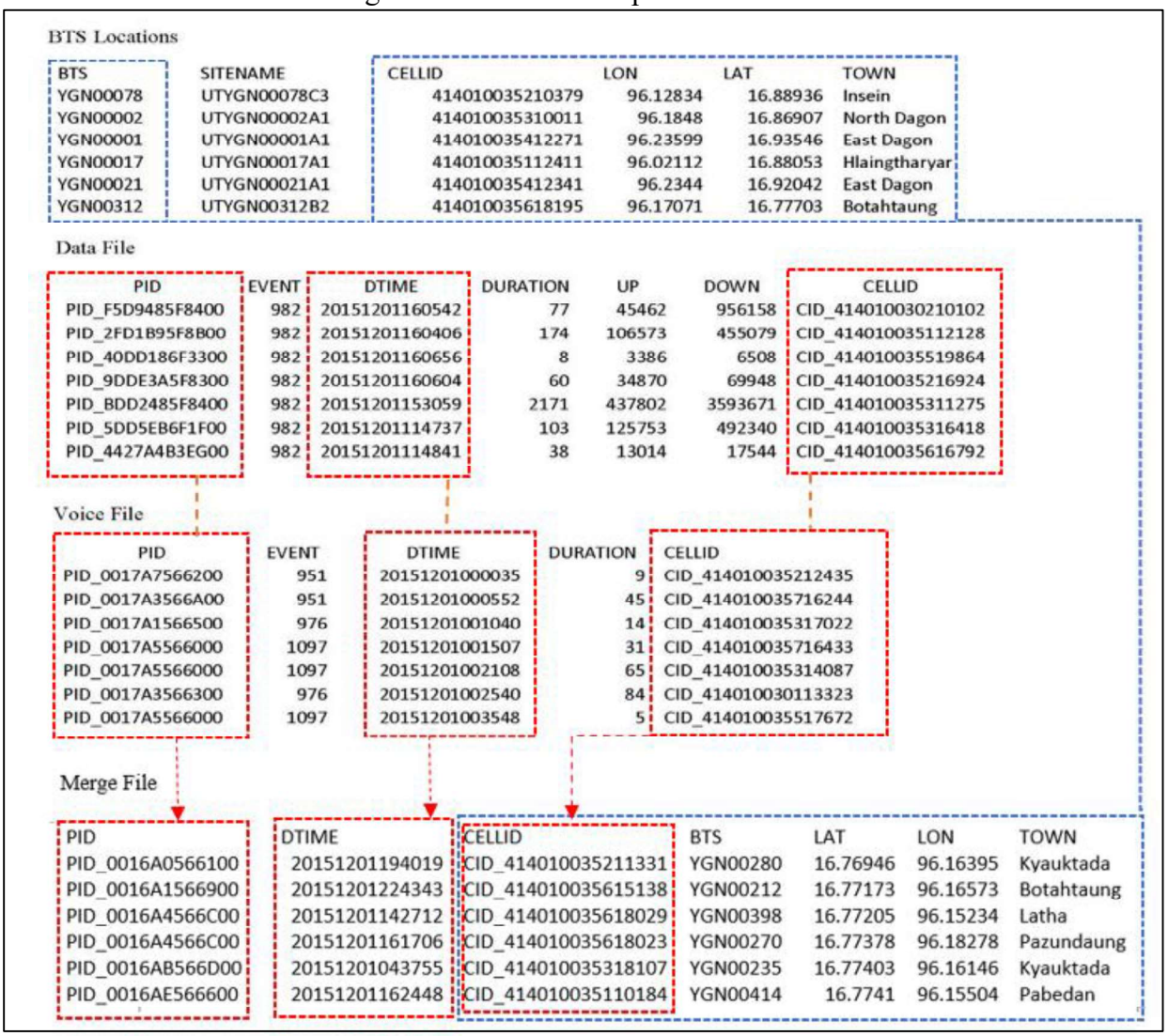


Figure 4. Formatted CDR data for voice and data with base transceiver station (BTS) locations

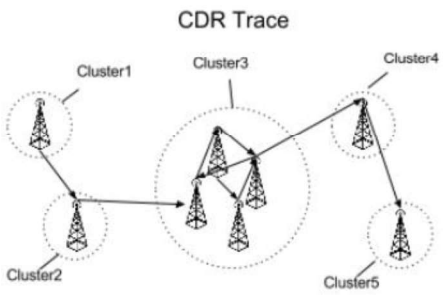
4.2.1.1 Noise Filtering

The connection of a mobile phone to the cell towers may hop between multiple towers even when the device stays at the same location. When a call cell tower hop occurs, a less crowded CELLID will serve the desired service. However, it creates a false trip, as the user has not changed his position. Such false movements can be identified by observing the fluctuation in people's locations. To solve the CELLID hopping problem, the noise filtering method (Yu Zheng et al., 2015) was used. This method is based on the observation of location changes with abnormally high speed. The key step is computing the travel speed of each point in a trajectory based on the consecutive time interval and travel distance between consecutive points. When cell tower A is recorded between multiple tower B records, and the speed between A and B is greater than the predetermined threshold (e.g., 200 km/h), such records are cut off.

4.2.2 Stay Point Detection

After preprocessing the CDR data, stationary stay points where the users engage in an activity were located. The stay point can be identified from the trajectory of consecutive cell phone records bounded by both temporal and spatial constraints. Two types of stay points occur in the trajectory. First, the records that show that a user remains immovable for longer than the temporal threshold of 10 min are considered as single stay points. The duration of stay was measured as the time difference between the first and last records.

Another type of stay point can be identified by clustering the records of nearby locations. In this case, the user might remain in the same location or keep moving around a small area. Therefore, we need to extract the consecutive points that are spatially close to the spatial constraint (less than 300 m), and temporal constraints (greater than 10 min). Then, those extracted points were considered as stay points. After that, the cluster of these stay points was conducted. We also found the centroids of these stay points to obtain a new set of stay points. The remaining stay points that would not be observed in any length of stay are called pass-by points.



CDR Log

PID	BTS	LAT	LON	Time
PID_0016A4566C00	YGN00282	16.77896	96.13801	6:49:58
PID_0016A4566C00	YGN00282	16.77896	96.13801	8:30:25
PID_0016A4566C00	YGN00282	16.77896	96.13801	9:59:08
PID_0016A4566C00	YGN00261	16.79305	96.14272	11:28:08
PID_0016A4566C00	YGN00295	16.80397	96.12353	12:59:08
.....				
P_n	BTS_n	LAT_n	LON_n	

Stay History

PID	LAT	LON	Time	Duration	Type
PID_0016A4566C00	16.77896	96.13801	9:59:08	11350	Stay
PID_0016A4566C00	16.79023	96.12629	13:38:30	3943	Stay
PID_0016A4566C00	16.77447	96.13971	15:33:37	1797	Stay
PID_0016A4566C00	16.77896	96.13801	16:35:21	20468	Stay

Figure 5. Stay point detection process

The pass-by points are filtered out, and it is assumed that the stay points are trip origins or destinations in which a trip is made. Figure 3 also shows the method of extracting a sequence of stay point locations from the raw CDR data.

4.2.3 Inferring Activity Locations (Trip Purposes)

People's mobility patterns are characterized by the frequency of visits to a particular location and therefore, people's activities can be identified using visit frequency. The activity types were labeled based on the specified rules (Xie et al., 2009; Huang et al., 2010; Phithakkitnukoon et al., 2010; Gong et al., 2015). The used labels are home, work, school, shopping, recreation, personal business, and others.

4.2.3.1 Identifying Home Location

Each user's home location was identified as the one with most frequent communication with the cell tower during the night time (from 10 pm to 7 am) on weekends and weekdays within a one-week period. After processing the home detection process, as shown in Figure 6, home towers are obtained for 1.5 million users. It is approximately 70% of the total phone users and approximately 37% of the total population. The expansion data from mobile phone users to study the area population will be discussed in Section 4.2.4.

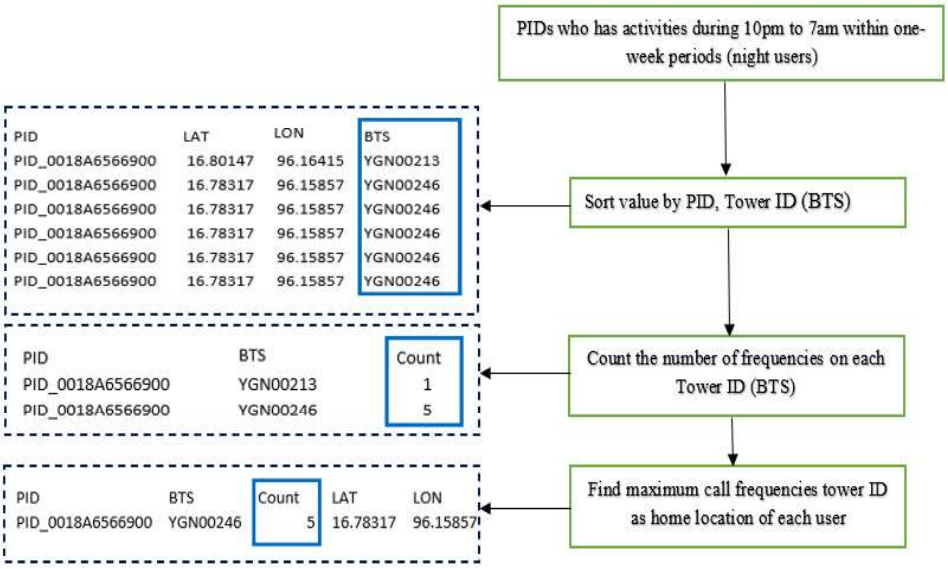


Figure 6. Home location identification

4.2.3.2 Identifying Work Location

For detection of work location, we identified the BTS of each user during working between 9 am to 6 pm, and then divided the accessed BTS of each user into two groups such as before-lunch group (9 am to 1 pm) and after-lunch group (1 pm to 6 pm). The common BTS in the two groups was then grouped. Subsequently, the work location was defined as the extracted common BTS to which the user travels the maximum number of times, max($d \times n$), where n is the number of times the user visited the cell tower during working hours, and d is the straight-line distance between the home location and the given location.

Suppose the distance from the identified work location to the home location is less than 0.5 km. In that case, stay region activity is identified as other activity instead of work, because work trips normally involve longer distance trips than short distance trips. The process is presented in Figure 7.

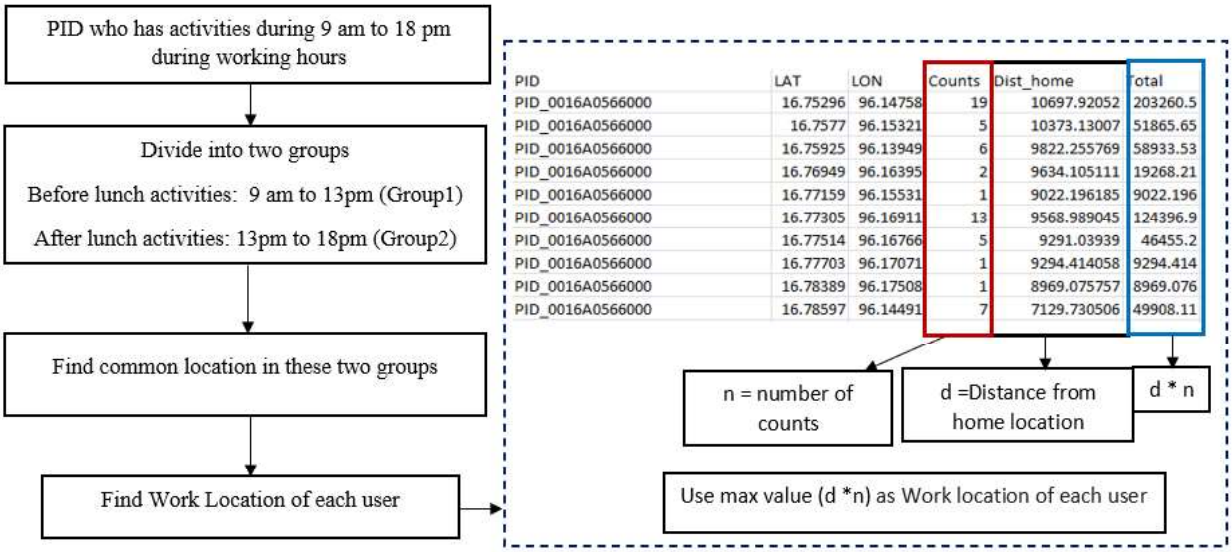


Figure 7. Work location identification

4.2.3.3 Identifying Other Location with Activities Types

In this study, four different human activities were considered. These activities were categorized into classes such as “school”, “shopping”, “social and recreational”, “personal business”, and “other”.

Labeling the types of activity based only on CDR data is very difficult because of the low spatial resolution. The land-use information within the catchment zone of a BTS is used to determine the likelihood of the corresponding activity. Here, the visit time interval and stay duration (Xie et al., 2009; Huang et al., 2010; Gong et al., 2015) are the key parameters. To detect the type of activities in each location, the study area was divided into Voronoi polygons for the network of BTS. Subsequently, a ranked list was prepared based on the probability of visits to the possible points by each user, and the possible type of activities was assigned to each location. Table 1 presents the aggregated activity categories with keywords used for the POI search. The distribution of POI has been presented in Figure 8.

Table 1. Aggregated activities and keywords used for POI
Source: (Phithakkitnukoon et al., 2010)

Activity category	Keywords used
School	University /School/ College/ Library
Social and recreational	Artwork/ Bar/Restaurant /Hotel/Rest area/Zoo/ Place of worship/ Monastery/Park/Gym/Playground
Shopping	Mall/Marketplace/Shop/Shopping Center/Super Market/Store/Plaza/Boutique/Automotive Shop/Food & Drink Shop/Bookstore/ etc.
Personal business	Bank /Clinic/ hospital/dentist clinic/Post-office/Embassy/Count house

Consider that one region has 13 social and recreational spots, eight schools, and three shopping spots. When users visit the target region, the probability of a trip for social and

recreational purposes is the highest if the trip properties (i.e., frequency, stay duration, and access time) match the corresponding properties listed in Table 1. If the trip properties of the social and recreational spots do not satisfy the trip-purpose condition, the trip for schooling will be analyzed. Even if the trip properties for schooling do not match, the third option will be analyzed and so forth. Then, it will be decided which type of activity would be most suitable for each location visited by the user.

Table 2. Rules for labeling activities based on spatiotemporal features

Character of the activity	Duration		Time interval
	t_{min}	t_{max}	
School	> 2 h	any	6:00–15:00
Social and recreational	1 h	5 h	any
Shopping	30 min	2 h	6:00–21:00
Personal business	30 min	any	7:00–21:00
Other	-	-	-

Table 3. Mapping of activity categories into activity types

Item	Aggregated activity category	Activity type
1	Home	All home activities
2	Work	All activities at work/job
3	School	Attending class, all other activities at school
4	Shopping	Routine shopping, shopping for many purchases or special items
5	Social and recreational	Outdoor and indoor recreation/entertainment, religious activities
6	Personal business	Personal business (visiting government office, attorney), household errands (bank, gas, etc.)
7	Other	Other unclassified activities

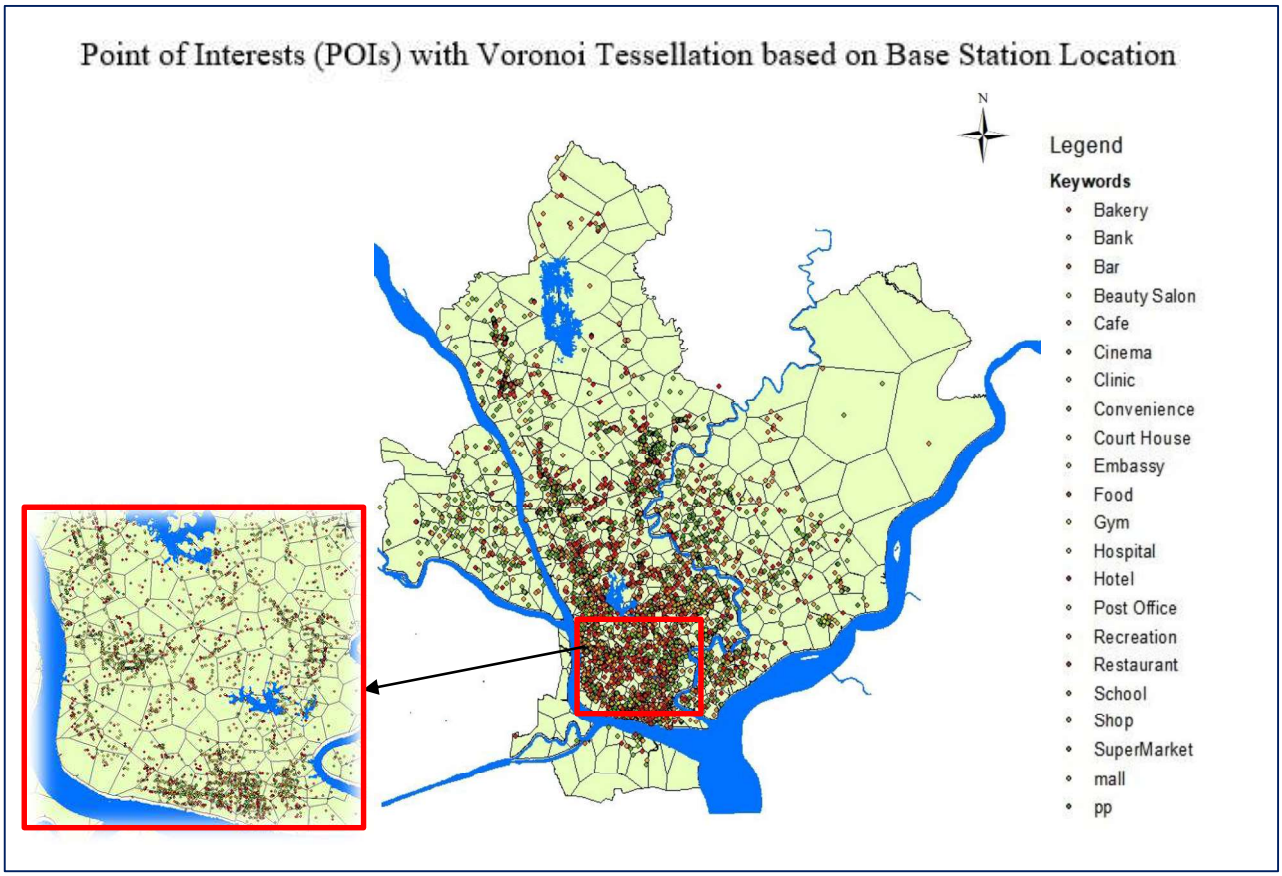


Figure 8. Point of Interest (POI) search results from Yangon open street map with Voronoi tessellation map

4.2.4 Data Expansion

For upscaling filtered mobile users to the total population of the study area, the 2014 Census population database was utilized. This database contains an aggregated population of the township. Therefore, we needed to disaggregate the census population into the base station population. Then, expansion factor, calculated as the ratio of the census population of each base station and the number of homestays identified in mobile phone data, is aggregated into each base station.

4.2.4 Determining Trips

The methodology for trip estimation is explained here. Stay location, activities, and time of day are combined to construct the user's origin–destination (O–D) trips. A trip is assumed for each mobile user when a trip is made between the two consecutive stay points of the user. Trip purpose is identified from the type of destination of stay location activities such as school trips, work trips, and shopping trips. The number of trips (flows) is determined as people's movement between activities that involve travel from the starting point in the origin region and ending in the destination region. To obtain the average daily O–D trips, each user's trips are multiplied by the expansion factors generated from the 2014 census data.

Table 4. Example of O-D trips with purpose for each user

PID	LAT	LON	Departure time	Arrival time	Origin	Destination	Purpose
PID_0016A4566C00	16.77896	96.13801	7:45:35	9:54:45	Home	Work	Work
PID_0016A4566C00	16.79023	96.12629	12:59:08	15:38:40	Work	Shopping	Shopping
PID_0016A4566C00	16.77447	96.13971	17:03:40	18:45:37	Shopping	Recreational	Recreational
PID_0016A4566C00	16.77896	96.13801	20:35:21	21:16:29	Recreational	Home	Home

5. RESULTS AND DISCUSSIONS

5.1 Trip Production and Attraction

After preprocessing the CDR data steps, we infer the user's home location and upscale the data to the population. These datasets represent the spatial distribution of population density aggregated in the traffic analysis zone of the study area. Trips are extracted from consecutive stay points from which trip production and attraction of each TAZ will be produced. Figure 9 illustrates the trip production and attraction for each TAZ with population density.

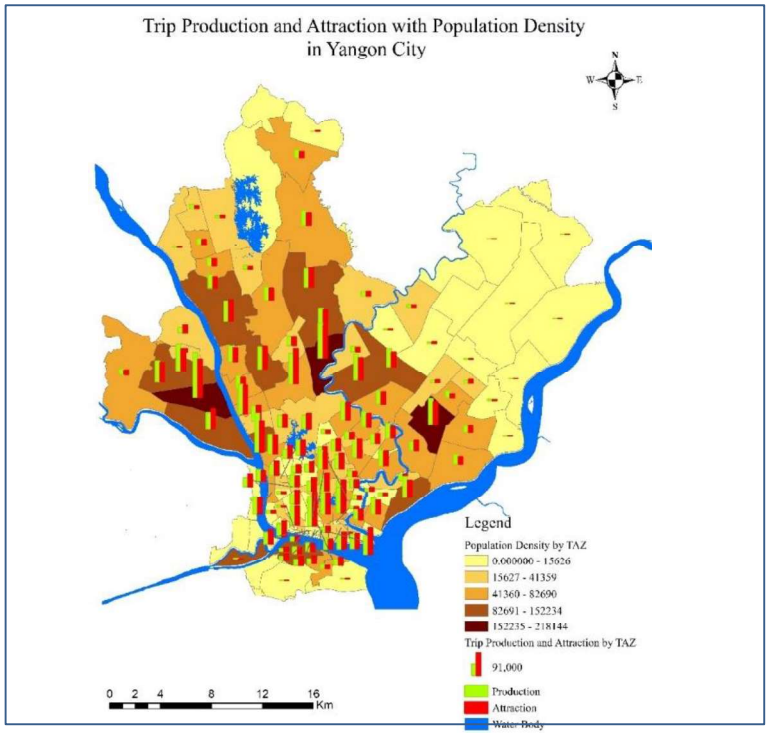
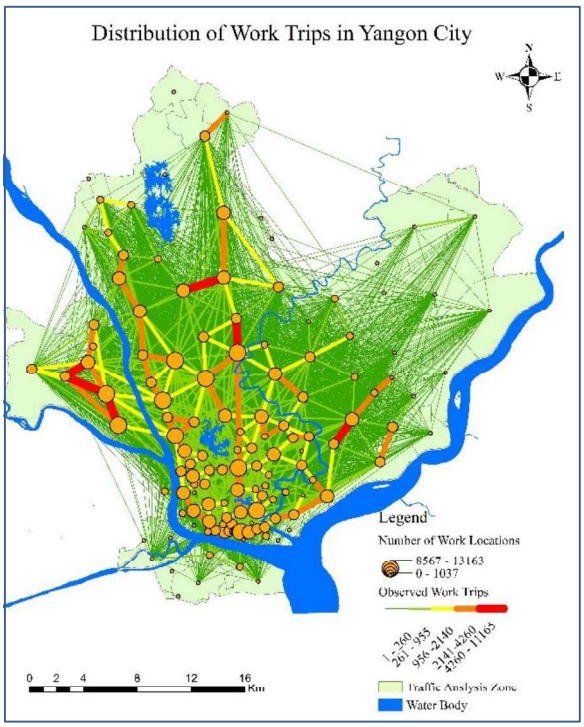


Figure 9. Example of O–D trips with purpose for each user

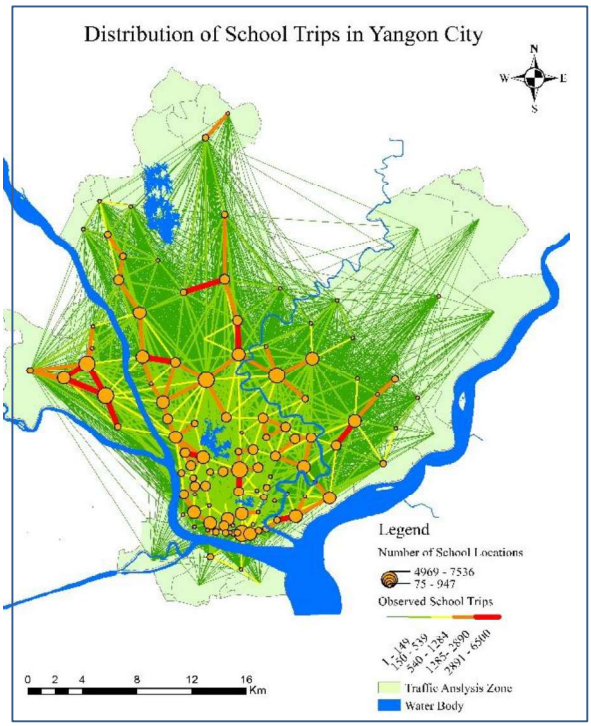
5.2 Trip Distribution of Activity Trips Based on CDR Data

After the trips are extracted, the travel demand between each pair of zones is computed. Figure 10 illustrates the spatial distribution of activity trip patterns for (a) work trips, (b) school trips, (c) shopping trips, (d) personal business trips, (e) home trips, and (f) social and recreational trips. These Figures show how the particular activity-type trips are distributed between different O–D pairs. This spatial distribution of travel demand is converted to aggregated trips between

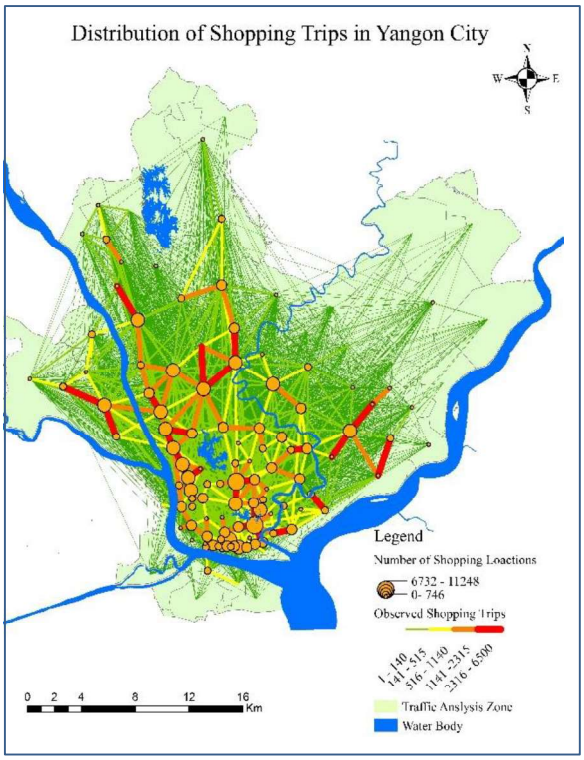
traffic analysis zones (TAZs) instead of trips between base stations. It can be seen from these figures that many people chose their destination's places based on the popularity of their corresponding activity place. Moreover, the relationship between the popularity of the activities and distinct trip distribution patterns can be found. This pattern implies a strong correlation between the urban activity of spatial distribution and destination choices.

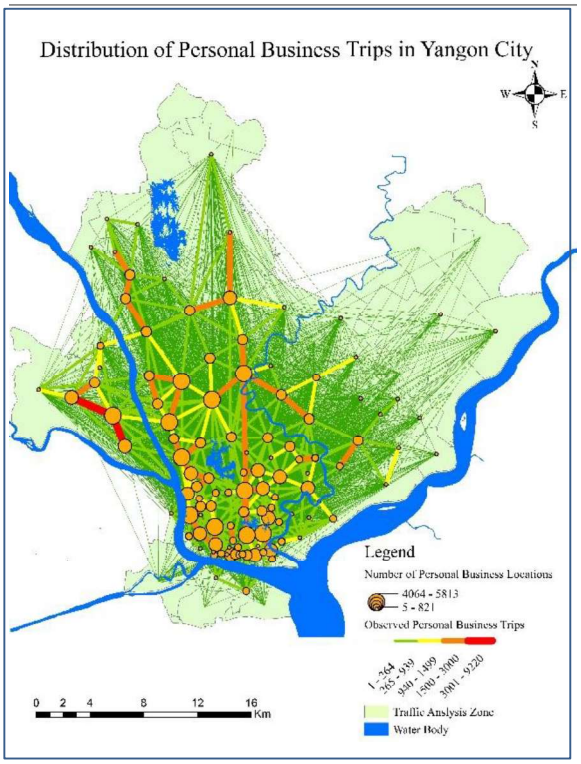


(a)

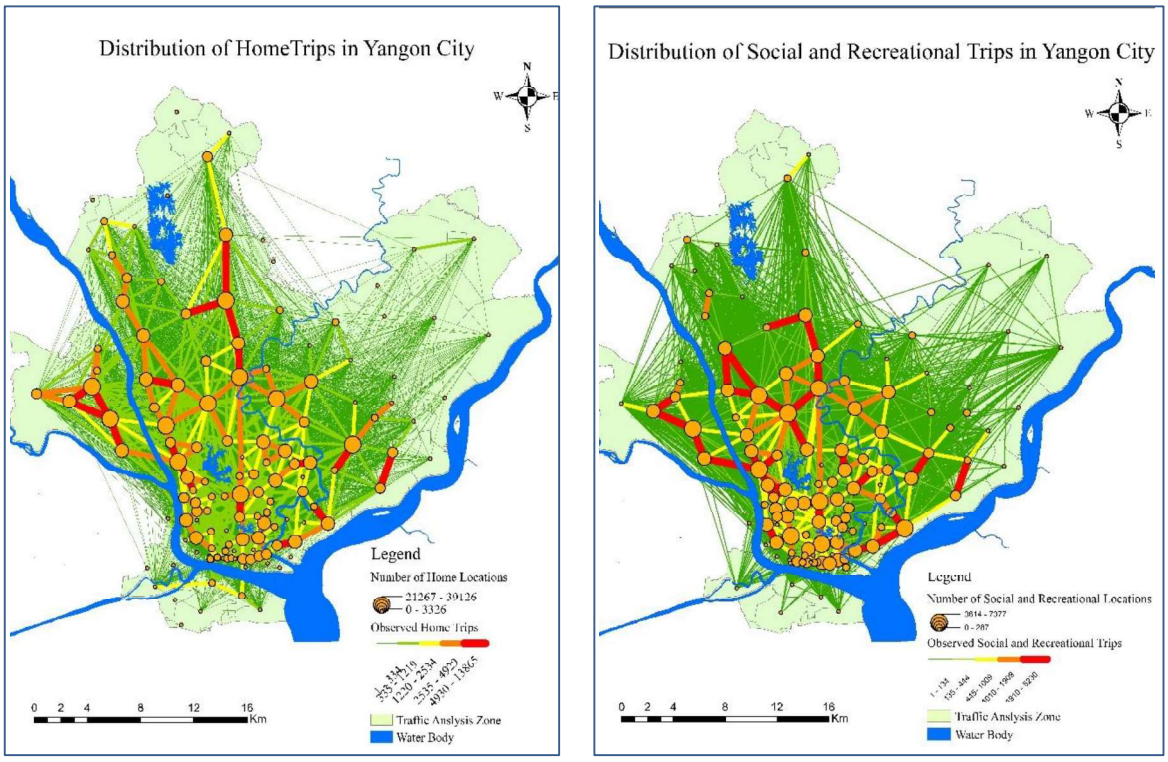


(b)





(c) (d)



(e) (f)

Figure 10. Spatial distribution of TAZ-pair CDR flow: (a) work trips; (b) school trips; (c) shopping trips; (d) personal business trips; (e) home trips; and (f) social and recreational trips

5.3 Distribution of Human Activity Patterns Based on CDR Data

After preprocessing and processing of the CDR data, trips are extracted from the O-D pairs and classified as activities with destination points. The activity trips are classified into seven trip purposes: school trips, work trips, home trips, shopping trips, social and recreational trips, personal business trips, and others. Each group of trips by purpose and the number of trips was extracted. Figure 11 illustrates the distribution of hourly activity arrival times for different purposes for the entire day.

From the figure 11, it can be seen that there is a slight movement of human activity between 12 am and 6 am. The peak volume occurs between 8 am and 10 am, and at 1 pm, 6 pm, and 8 pm. Most people prepare to go to their work in the morning, and hence the hourly pattern of work trip activity gradually climbs to the highest peak during the morning peak period (8–10 am). It falls slightly from 10 am to 12 pm and increases slightly during the noon peak period. Subsequently, it slows down dramatically in the following hours of the day. It can be seen that the work activity level in the morning peak hours is higher than that in the evening peak period. The school activity occurs with the highest volume during the morning peak period; afterward, it decreases slightly from 9 am to 4 pm. It can be inferred that work trips and school trips are primarily responsible for the morning peaks

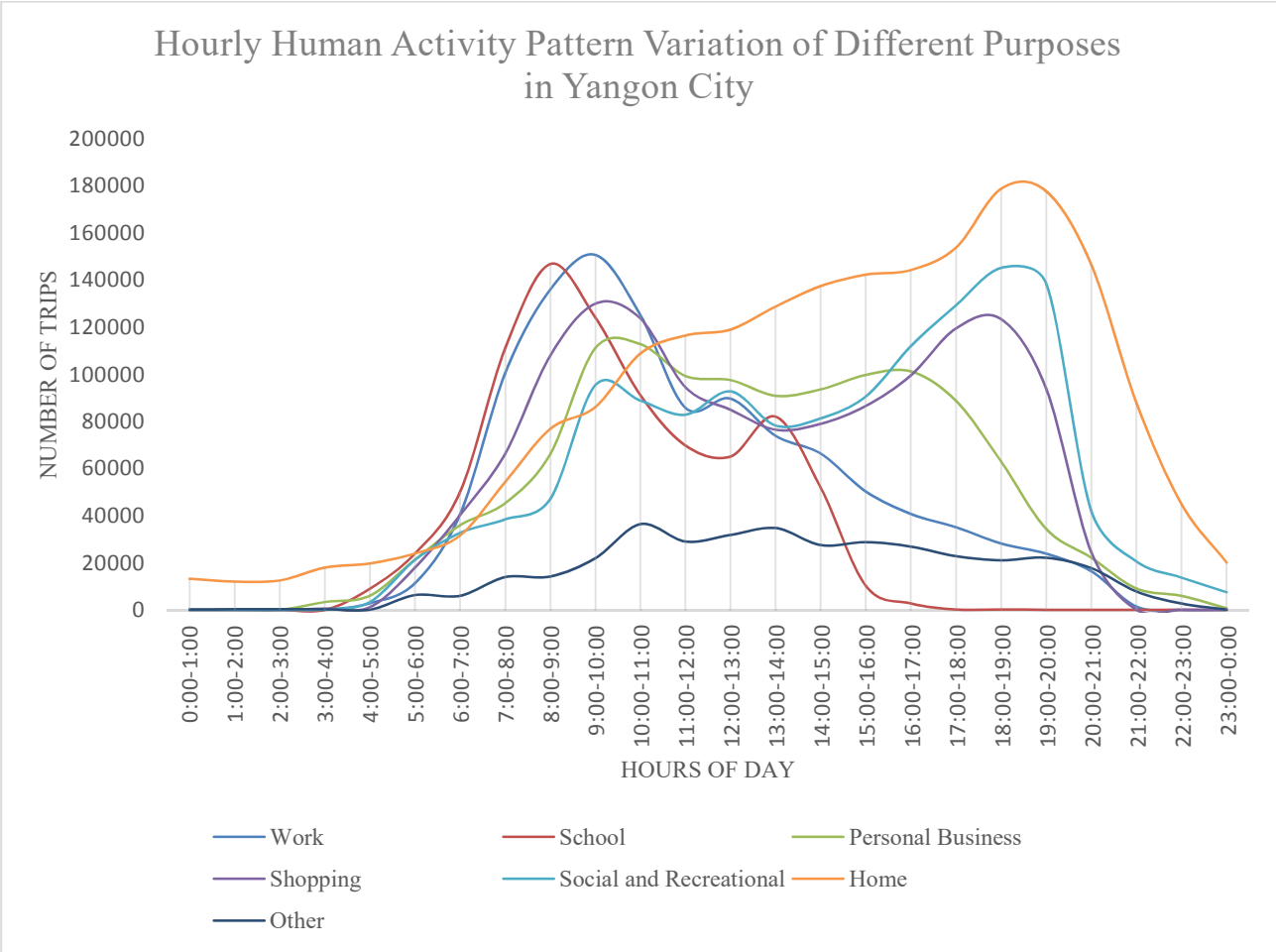


Figure 11. Distribution of activity start time with different purposes

.

The highest volume of personal business occurs during the morning peak; it remains high between 11:00 am and 4:00 pm, and then falls steadily during the night. Moreover, hourly human activity variations in shopping and recreation have similar trends. It increases slightly, and a higher volume occurs during the morning and evening peak periods. Most people return from their activity places to their homes in the evening, and hence the highest peak in the activity level is observed during the evening peak periods. Furthermore, the other activities fluctuated slightly throughout the day. Overall, it can be seen that the total number of trips during the evening periods is higher than that during the morning peak periods.

Therefore, it is found that school and work trips are the highest during the morning peak periods, and home, shopping, and recreational trips are the highest during the evening peak periods.

5.4 Validation by JICA Person Trip Survey (YUTRA, 2014)

After identifying trips by purpose using the CDR data, the results were validated by the YUTRA-JICA person trip ground truth survey (YUTRA, 2014). The comparison of trip purposes by the JICA Person Trip Survey (YUTRA Survey 2014) and those based on CDR data (2015) is shown in Figure 12.

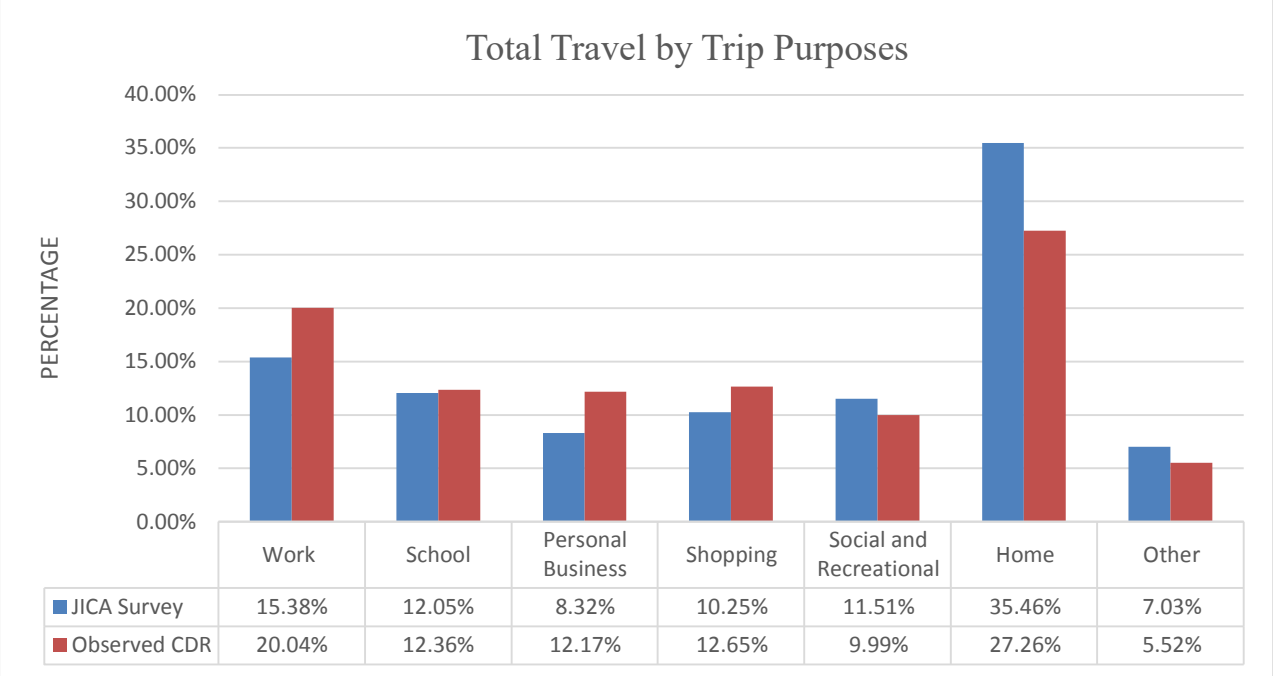


Figure 12. Comparison of trips by purpose in Yangon City

The trips by purpose identified using the CDR data were 13.02 %, 10.15%, 14.59%, 16.54%, 16%, 24.82%, and 4.53% for work, school, personal business, shopping, social and recreational, home, and other trips, respectively, in the study area. From the JICA Survey, 8.78%, 12.81 %, 13.35%, 18.16%, 14.76%, 16.26%, and 15.88 % of the trips were made for work, school, personal business, shopping, social and recreational, home, and other trips, respectively. By comparing these two data sources, it can be found that the results are approximately the same, but some differences occurred owing to the absence of BTS towers in some traffic analysis zones, dead BTS of the study area, and limitations of the CDR data.

5.4 Limitations of CDR and Other Data

Applying mobile phone record data for the transportation sector can benefit future planning. However, these records have some limitations. Although there are three telecommunication operators in Myanmar, such as Ooredoo, Telenor, and MPT services, only MPT CDR data are available, while there are limitations in the data from the other operators. While the CDR data were used in this research, the details can be derived by tracing the trajectory location of each user. However, there is a lack of continuity of mobility traces, because users do not use their mobile phone data when they move from one place to another. The working location is estimated based on working in the daytime, but the user may be working during normal working time or taking overtime. Sometimes, the embedded individual activity is correlated with land use, visitation time interval, and duration, and hence potential and temporal activity bias would occur because of the mixed land-use effect. Finally, the visitation time interval and duration were based on the travel behavior characteristics of the study area, which might be a subject future research.

Although there were 132 traffic analysis zones in the study area, only 122 traffic analysis zones were analyzed to extract trips from the O-D pair. The remaining ten traffic analysis zones were not supported by BTS towers owing to dead 57 BTS towers during the study period.

Detailed land-use databases and data on POI were not available consistently throughout the study area owing to constraints on search limits and capabilities, which may introduce errors.

6. CONCLUSION

This study identified individual human activity patterns and origin-destination trips by purposes based on activity spatiotemporal features of aggregate with traffic analysis zones using CDR data; the results were validated with the JICA data. Trip purposes were identified by integrating land-use data with CDR data. The results were validated as there was only a small difference between the results from the CDR data and the JICA data.

The study has three major findings. Firstly, this analysis revealed that school and work trips were the highest during morning peak hours, and that home, social and recreational trips, and shopping trips were the highest during the evening peak hours in Yangon's study area. Second, we found that in most trips, the destination was selected based on the popularity of that place. Lastly, the spatiotemporal distribution of a trip and the infrastructure attributes of those places are closely linked. This pattern implies a strong correlation between people's activity participation and destination choices.

This study demonstrates that CDR data can be effectively used to infer embedded human activity patterns relevant to transportation planning applications. Furthermore, these output results can be applied to the trip generation and distribution steps of traditional four-step travel demand models. Therefore, these data can be directly applied to the development of transportation decisions and policies to build a sustainable future. This study will improve existing urban trip models and applications related to the mitigation of congestion in Yangon city, traffic planning, and public facility management.

In further studies, the spatial distribution of activity patterns and their dynamics will be analyzed to gain insight into crowd behavior for future planning and operation of the underlying transport networks.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Myanmar Posts and Telecommunications (MPT) for providing the CDR of one-week data for this study.

REFERENCES

Acker, V.V., Goodwin, P., Witlox, F. (2016) Key research themes on travel behavior, lifestyle, and sustainable urban mobility. *Int. J. Sustainable Transp*. 10, 25-32.

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(B), 240–250. ISSN0968-090X. https://doi.org/10.1016/j.trc.2015.02.018.

Batran, M., Mejia, M., Kanasugi, H., Sekimoto, Y., Shibasaki, R. (2018) Inferencing Human Spatiotemporal Mobility in Greater Maputo via Mobile Phone Big Data Mining. *ISPRS International Journal of Geo-Information*, 7(7), 259. https://doi.org/10.3390/ijgi7070259.

Boarnet, M.G., Hsu, H.P., (2015) The gender gap in non-work travel: The relative roles of income earning potential and land use. *J. Urban Econ*. 86, 111-127.

Calabrese F, Diao M, Di Lorenzo G, Lorenzo, D.G., Ferreira, J., Ratti, C.. (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301–313.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285-299. https://doi.org/10.1016/j.trc.2016.04.005.

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. *Transp. Res. Rec*, 2526, 126-135.

Dash, M., Koo, K.K., Decraene, J., Yap, G.E., Wu, W., Gomes, J.B., Li, X. (2015) CDR-To-MoVis: Developing a Mobility Visualization System from CDR data. 2015 IEEE 31st International Conference on Data Engineering. DOI: 10.1109/icde.2015.7113399.

Demissie, M.G., Phithakkitnukoon, S., Kattan, L. (2018) Trip Distribution Modeling Using Mobile Phone Data: Emphasis on Intra-Zonal Trips. Paper presented at the IEEE Transactions on Intelligent Transportation Systems, 20(7), 2605–2617. https://doi.org/10.1109/tits.2018.2868468.

Demissie, M.G., Phithakkitnukoon, S., Kattan, L., Farhan, A. (2019) Understanding Human Mobility Patterns in a Developing Country Using Mobile Phone Data. *Data Science Journal*, 18(1), 1-13. DOI: https://doi.org/10.5334/dsj-2019-001.

Diao, M., Zhu, Y., Ferreira, J., (2015) Inferring individual daily activities from mobile phone traces: A Boston example. *Environment and Planning B: Planning and Design*, 43(5), 920–940. DOI: https://doi.org/10.1177/0265813515600896

Eagle, N., Macy, M., Claxton, R. (2010) Network diversity and economic development. *Science*, 328(5981), 1029-1031.

Gong, L., Liu, X., Wu, L., Liu, Y., (2015) Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103–114.

Guan, C., Song, J., Keith, M., Akiyama, Y., Shibasaki, R., Satoe, T., (2020) Delineating urban park catchment areas using mobile phone data: A case study of Tokyo. *Computers, Environment and Urban Systems*, 81, 101474. https://doi.org/10.1016/j.compenvurbsys.2020.101474.

Gundlegård, D., Rydergren, C., Breyer, N., Rajna, B. (2016) Travel demand estimation and network assignment based on cellular network data. *Computer Communications*, 95, 29-42. https://doi.org/10.1016/j.comcom.2016.04.015.

He, S., (2011) The effect of school quality and residential environment on mode choice of school trips. *Transp. Res. Rec*. 2213, 96-104. https://doi.org/10.3141/2213-13.

Huang, L., Li, Q., Yue, Y., (2010) Activity identification from GPS trajectories using spatial temporal POIs' attractiveness, Paper presented at the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. ACM, New York, pp. 27–30.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., Rowland, J., Varshavsky, A. (2011) Ranges of Human Mobility in Los Angeles and New York, Paper presented at the 2011 IEEE International Conference on Pervasive Computing and Communication Workshops, pp. 83-93.

Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C., (2014) Development of origin-destination matrices using mobile phone call data. *Transp. Res. Part C*, 40, 63-74.

Japan International Corporation Agency (2014) Project for the Comprehensive Urban Transport Plan of the Greater Yangon YUTRA. Accessed from https://openjicareport.jica.go.jp/pdf/12182754.pdf (December 2020)

Jiang, S., Ferreira, J., González, 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. https://doi.org/10.1109/tbdata.2016.2631141.

Kung, S.K., Greco, K., Sobolevsky, S., Ratti, C. (2014) Exploring universal patterns in human home-work commuting from mobile phone data. *Plos One*, 9(6). https://doi.org/10.1371/journal.pone.0096180.

Kyaing, D., Lwin, K., Sekimoto, Y. (2017) Human Mobility Patterns for Different Regions in Myanmar Based on CDRs Data. IPTEK J. Proc. Ser. 2017. DOI: 10.12962/j23546026.y2017i6.3271

Lane N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T. (2010) A survey of mobile phone sensing. *IEEE Communications Magazine*. 48(9), 140-150. doi: 10.1109/MCOM.2010.5560598.

Lee, K., You, S.Y., Eom, J.K, Song, J., Min, J.H. (2018) Urban spatiotemporal analysis using mobile phone data: Case study of medium- and large-sized Korean cities. *Habitat International*, 73, 6-15.

Louail, T., Lenormand, M., Picornell, M., Cantú, G.O., Herranz, R., Frias-Martinez, Ramasco, J.J, Barthelemy, M. (2015) Uncovering the spatial structure of mobility networks. *Nature Communications*, 6, 6007.

Lu, X., Bengtsson, L., Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29), 11576-11581. DOI:10.1073/pnas.1203882109.

Lwin, K.K., Sekimoto, Y. Takeuchi,W. (2018)Estimation of Hourly Link Population and Flow Directions from Mobile CDR, International Journal of Geo-Information 7 (11) 449-461. doi:10.3390/ijgi7110449

Phithakkitnukoon, S., Horanont, T., Lorenzo, G., Shibasaki, R., Ratti, C. (2010) Activity-aware Map: Identifying Human Daily Activity Pattern using Mobile Phone Data**.** MIT-Senseable City Lab, Boston.

Shen, L., Stopher, P.R.,(2014) Review of GPS travel survey and GPS data-processing methods. *Transp. Rev*. 34, 316-334.

Song, C., Qu, Z., Blumm, N., Barabasi, A. (2010). Limits of Predictability in Human Mobility. *Science*, 327(5968), 1018-1021. DOI:10.1126/science.1177170.

Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., González, M.C. (2012) Understanding road usage patterns in urban areas. *Sci. Rep*, 2.

Wesolowski, A., Eagle, N., Tatem, A., Smith, D., Noor, A., Snow, R., Buckee, C. (2012) Quantifying the Impact of Human Mobility on Malaria. *Science*, 338(6104), 267-270.

Xie, K., Deng, K., Zhou, X., (2009) From trajectories to activities: a Spatio-temporal join approach. Paper presented at the 2009 International Workshop on Location Based Social Networks. ACM, New York, pp. 25–32.

Zheng, Y. (2015) Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol*, 6(3), 1-41. DOI: http://dx.doi.org/10.1145/2743025