Inference of Human Spatiotemporal Mobility in Greater Maputo by Mobile Phone Big Data Mining

Mohammed Batran[1], Mariano Gregorio Mejia[1], Yoshihide Sekimoto[1], Ryosuke Shibasaki[2]
[1]Institute of Industrial Science (IIS), the University of Tokyo
(batran, mgbmejia, Sekimoto)@iis.u-tokyo.ac.jp
[2]Center for Spatial Information Science, The University of Tokyo
shiba@csis.u-tokyo.ac.jp

Abstract

This study demonstrates how big data sources, particularly, mobile phone call detail record (CDR) data can be used to infer human spatiotemporal mobility patterns that would be valuable for urban and transportation planning purposes. The daily travel behavior of people have been commonly derived from traditional data collection methods such as person-trip interview surveys, which are generally costly and difficult to implement especially in developing cities with a high population and limited resources. The ubiquitous massive and passive CDR data provides researchers an opportunity to innovate alternative methods that are inexpensive and less complex while maintaining an acceptable level of accuracy as that of traditional ones. This study aims to capture the daily mobility of people living in Greater Maputo, and in turn capture human activity locations of interest. Accordingly, we propose a method based on proven techniques to extract origin-destination (OD) trips from raw CDR data of 3.4 million mobile phone users in a 12-day period, and scale the processed data to represent the mobility of the actual population in different time frames (weekday and weekend). The output mainly include trip generation and attraction maps of Greater Maputo that would be especially useful for planners and policymakers. In order to evaluate the performance of the proposed method, the results are validated with actual survey data from the Japan International Cooperation Agency.

1 Introduction

The quality of urban life and a city's economic growth may lie in how well its urban spaces and transportation infrastructure have been planned for and developed. In urban and transportation planning, it is critical to consider how to efficiently and effectively move people and goods in the city from their points of origin to their respective destinations by understanding how people conduct their activities daily and with that their travel behavior. Traditional methods persist to collect such needed travel and activity behavior data by means of household and person-trip interview surveys, which are usually limited to a smaller sample size and involves a larger scale of deployment, and at that, are resource intensive in terms of time, cost, and labor. In light of advances in digital and sensing technologies that acquire big data, researchers in urban and transportation planning-related fields have been coming with new approaches that utilize such as a potentially better alternative to traditional ones. One approach involves big data analytics utilizing call detail records (CDRs)—which are digital records passively produced and collected by telecommunications equipment for each instance of mobile phone communication usage (i.e., voice calls, short message service (SMS), and internet service). Recent studies have demonstrated the application of CDR data analytics to infer travel behavior and understand human mobility.

Zin *et al.* [2018] estimated people movement in Yangon within a limited time frame separately for weekdays and weekends based on CDR data. The origins and destinations of the trips are taken based on traffic analysis zones corresponding to the cellular tower in which the record was made. Jiang *et al.* [2015] utilized CDR in Singapore to examine human travels in an activity-based approach that focuses on patterns of tours and trip-chaining behavior in daily mobility networks. The authors developed an integrated pipeline, which includes parsing, filtering and expanding massive and passive raw CDR data and extracting meaningful mobility patterns from them that can be directly used for urban and transportation planning purposes. Phithakkitnukoon *et al.* [2010] developed an *activity-aware map* that describes the probable daily activities of people (during weekdays) for specified areas based on CDR data. Their results showed a strong correlation in daily activity patterns within the group of people who share a common work area's profile. Meanwhile, Wang *et al.* [2010] proposed a method to infer transportation mode based on travel time extracted from CDR data in the city of Boston. Although CDR data is coarse-grained, the authors' method demonstrated acceptable accuracy as that using fine-grained data but with the advantage of being low cost and suitable for statistical analysis on transportation modes of a large population.

This study demonstrates the utilization of big mobile phone data to extract meaningful information for understanding the daily mobility of people, in this case, in Greater Maputo—useful for urban and transportation planning purposes, and keeping in mind the advantages of our proposed method with traditional data collection methods. We validate our results with actual survey data to determine our proposed method's performance.

The rest of this paper is presented as follows: Section 2 explains in detail the study area and the data used, which include the CDR data for trip estimation, the High Resolution Settlement Layer (HRSL) population data for expanding the user sample to population, and the JICA survey data for validation purpose. Section 3 explains step-by-step the proposed methodology, from preparing the study area to upscaling the sample to represent the actual population. Section 4 presents the results of the method, and its validation. Lastly, Section 5 presents our conclusion of our method, including its potential for actual application in urban and transportation planning and development.

2 Study Area and Data

2.1 Study Area

The study area is Mozambique's Greater Maputo metropolitan area—consisting of the capital city Maputo, Matola City, Boane City, and Marracuene District. In recent years industrial and residential development and the growing urban population have spread from Maputo City—the country's political and industrial center—to its neighboring areas of Matola, Boane, Marracuene, creating the 120,767-ha Greater Maputo metropolitan area, as shown in Figure 1 [JICA, 2014]. According to JICA's forecast for the medium term from years, 2012 to 2035, the population of Greater Maputo is expected to increase from 2.2 million to 3.7 million, and its economy growing 2.3 times in terms of gross domestic product (GDP) per capita. With urban and economic development, Greater Maputo has seen more movement of people and goods, and with it, worsening traffic conditions in its underdeveloped road network. The number of daily person trips was estimated to increase more than double, from 3.1 million trips/day in 2012 to 6.5 million trips/day in 2035, with car ownership increasing 1.5 times for the same medium-term period [JICA, 2014]. These rapidly growing development indicators point out the urgent need to formulate a comprehensive master plan that would facilitate implementation and improvement of Greater Maputo's public transport infrastructure and road network [JICA, 2014].

2.2 Call Detail Record Data

This study used mobile phone call detail record (CDR) data collected for a 12-day period, i.e., 1st to 12th of March 2016, from a major mobile network operator (MNO) in Mozambique. The raw CDR dataset contains a total of 393 million mobile phone usage records from 3.4 million anonymous subscribers nationwide of the MNO. As mentioned above, there are three main types of mobile phone usage considered, i.e., 1) making/receiving a voice call, 2) sending/receiving a text message by SMS, and 3) using data or internet service. Due to privacy issues, all personal information in the CDR that may reveal the subscriber's identity have been anonymized by the MNO prior to its distribution. The relevant information contained in the CDR data include the user ID, timestamp and type of mobile phone usage, and ID and location of the recording cellular tower.

Figure 2 shows the temporal distribution (daily and hourly) of the raw CDR dataset. It should be noted that the 12-day observation period consist of 9 weekdays and 3 weekends. The daily distribution shows a consistent number of recorded mobile phone usage for all days, with the exception of one day, i.e., 3rd of March 2016 (Wednesday), which has a relatively low number of records (minimum value). Meanwhile, it is interesting to see that the hourly distribution shows low mobile phone usage records between midnight and 5:00 AM, the time period when most of the population is expected to be sleeping. On the other hand, the most active period is between 6:00 PM and 8:00 PM. The trip identification efficiency would rely on the number of mobile phone records, which are relatively high from 8:00 AM until 11:00 PM based on the hourly distribution. It is expected that it is within this time frame that most trips occur—particularly people traveling between their home and workplace or some other place, and would statistically provide better trip estimates.

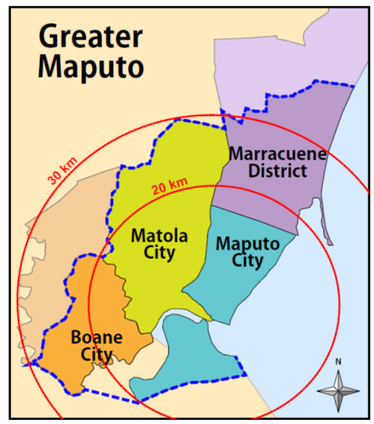


Figure 1. Map of Greater Maputo metropolitan area (study area) [JICA, 2014]

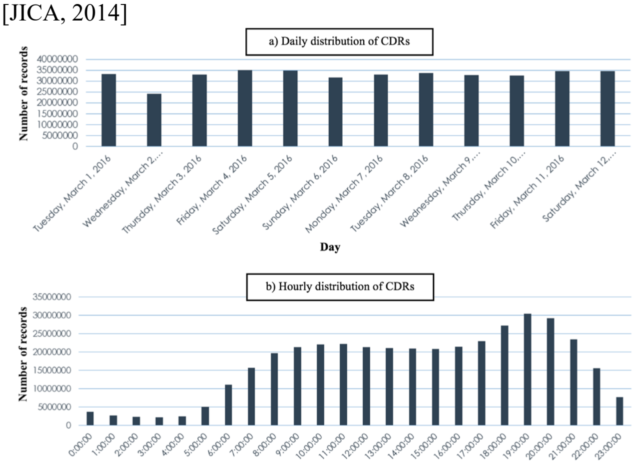


Figure 2. Temporal distribution of the CDR dataset: a) daily, and b) hourly

2.3 Population Data from the High Resolution Settlement Layer

This study used Greater Maputo's population data as obtained from the High Resolution Settlement Layer (HRSL) developed by the Facebook Connectivity Lab and Center for International Earth Science Information Network (CIESIN). The HSRL developed for Mozambique, as shown in Figure 3, provides estimates of human population distribution at a resolution of 1 arc-second (approximately 30 m) for the year 2015 based on recent census data and high-resolution satellite imagery from DigitalGlobe [Facebook and CIESIN, 2016]. Detailed information about the HRSL can be found from the website of CIESIN. Accordingly, the HRSL population distribution is used to estimate the population at the cellular tower zone level, which is represented by Voronoi zones, as discussed in Section 3.1. The estimated total population in the study area is at 2,661,832. Statistics on the resulting population distribution in the study area are summarized in Table 1.

Table 1. Population distribution in Greater Maputo

Min	Max	Mean	Median
62	45,261	10,277	6,781

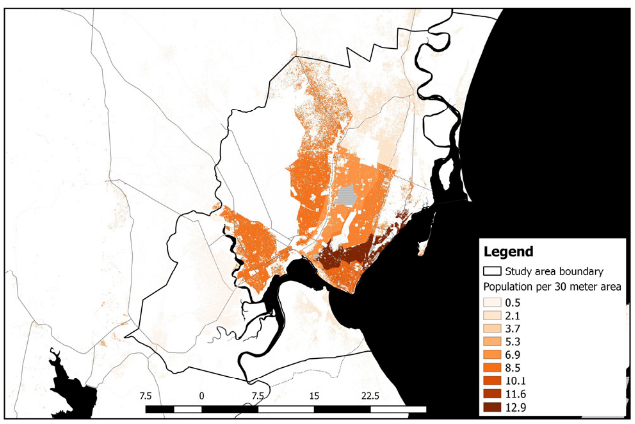


Figure 3. HRSL of Greater Maputo

2.4 JICA Survey Data for Results Validation

This study also uses person-trip survey data obtained from JICA's Comprehensive Urban Transport Master Plan for the Greater Maputo project [JICA, 2014]. The JICA survey sampled a total of 38,216 persons over the age of six from 9,983 sampled households, producing a total of 65,168 trips in one day. In the classical four-step demand forecasting model, the first step which is trip generation estimates the number of trips originating from and attracted to traffic analysis zones (TAZs) that are defined based on socio-economic, demographic and land use attributes of the cordoned area [McNally, 2007]. For the validation of our results, we consider three TAZ levels based on the zoning levels of the Greater Maputo metropolitan area in the JICA report, i.e., A TAZ, B TAZ, and C TAZ, as shown in Figure 4. The TAZs of C TAZ level were identified for the JICA survey and for transport modeling purposes, and they correspond to the administrative boundaries of the "*bairro*", where census data is available. While few *bairros* were consolidated to form bigger TAZs for B TAZ level, and further consolidation to four big TAZs for C TAZ level. Table 2 shows a statistical summary of the three TAZ levels.

Figure 4. Division of TAZs for each TAZ level

Table 2. Summary of TAZ levels

Zone Level	No. of TAZs	Min (km^2)	Max (km^2)	Mean (km^2)	Median (km^2)
C TAZ	170	0.03	95.20	7.10	1.08
B TAZ	40	0.81	305.13	30.21	9.65
A TAZ	4	252.61	381.09	302.17	287.49

3 Methodology

In extracting the origin-destination (OD) trips from raw CDR data, and scaling them to represent the mobility of the actual population of Greater Maputo in different time frames (weekday and weekend), we propose a method that incorporates proven techniques from previous research [Jiang *et al.,* 2017; Schneider *et al.,* 2013; Zin *et al.,* 2018]. Our method involves: 1) Voronoi tessellation of the study area, 2) estimation of the home location of subscribers, 3) filtering of valid user-days of the sample, 4) extraction of mobility/OD trips of the filtered sample, and 5) application of two types of magnification factors, one to upscale the user sample to represent the actual population in each zone, and the other one to normalize the user sample to one observation day.

3.1 Voronoi Tessellation of the Study Area

There are 259 cellular towers in the study area, which are spatially distributed in relation to the distribution of population density. The density distribution of cellular towers, and correspondingly the mobile phone network coverage area increases toward the city center and central business district. Generally, there is overlapping of coverage areas of neighboring cellular towers, which should be considered in order to appropriately represent the locational boundaries of each tower. Accordingly, a centroidal Voronoi diagram of the cellular tower network in the study area was developed, as shown in Figure 5. Each Voronoi tessellation approximately represents the mobile phone network coverage, and at that, the area coverage of each cellular tower. Table 3 shows a summary of the Voronoi tessellations in the study area. The minimum area coverage is at 0.01 km^2, and the maximum at 241.21 km^2, while the mean and median are at 8.15 km^2 and 1.81 km^2, respectively.

Table 3. Statistical summary of Voronoi zones in the study area

Zone Source	Number of Towers	Min (km^2)	Max (km^2)	Mean (km^2)	Median (km^2)
CDR Voronoi	259	0.01	241.21	8.15	1.83

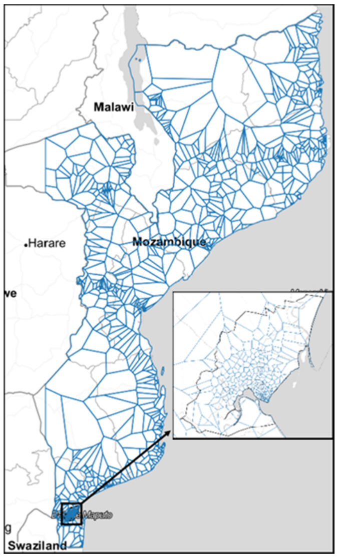


Figure 5. Voronoi diagram of the study area (magnified area), representing the mobile network coverage of each cellular tower

3.2 Home Location Estimation

There are several reasons why there is a need to identify the home location of the subscribers [Jiang *et al.,* 2017]. Firstly it would be needed when combining the CDR data and the population data from the HRSL for upscaling the user sample to represent the actual population. Secondly, we consider only the users living within the study area, such that, if a user is found to have a home location outside the study area then they are excluded from the sample. Thirdly, the environment and attributes of people's home, such as land use, affects their travel behavior and activities [Cervero and Murakami, 2010; Zergas 2004]. This is important to understand people's mobility and travel behavior as affected by the space or environment that surrounds them, especially in an urban planning and development perspective [Jiang *et al.,* 2017].

Accordingly, we estimate the most probable location of a subscriber's home based on their mobile phone usage records. It is expected that most people are at home at night and during the weekends, rather than during working hours on weekdays. The home location of subscribers is estimated based on the frequency of recorded mobile phone usage at cellular towers (corresponding to the Voronoi zones) at these times. The "night time" considered for this study is between 8:00 PM and 6:00 AM, since this time period was found out to be when most of the people in Greater Maputo are at home according to the JICA household survey. Also, most of the population reside outward of the city center in Maputo, in the outskirts of the study area, and therefore the cellular towers (Voronoi zones) located in these areas were considered as home locations of the subscribers. Other subscribers with home locations estimated outside of the study area were excluded from the sample. Therefore, from our home estimation location, we found 1,279,291 subscribers living within the study area. This corresponds to 48% of the total population of Greater Maputo estimated using HRSL (2,661,832 people). Figure 6 shows the distribution of subscribers living in the study area resulting from the home location estimation.

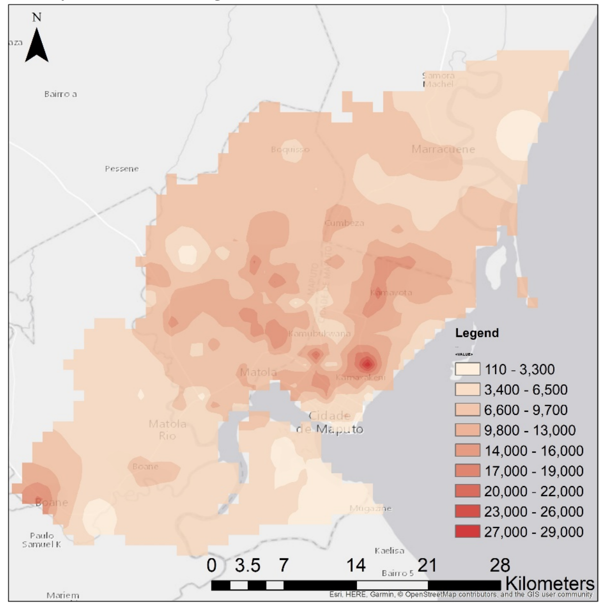


Figure 6. Subscriber distribution in the study area

3.3 Filtering Valid User-Days

According to Jiang *et al.* [2017], the advantages of CDR data are its longer sample period and larger sample size as opposed to traditional survey data, and its disadvantage is its sparseness. As such, it is important to consider user-days with much mobile phone activity or usage. According to previous research [Jiang *et al.,* 2017; Schneider *et al.,* 2013], extracting individual mobility patterns from CDR data can be regarded statistically consistent and comparable with that of traditional survey data given that a certain threshold of daily mobile phone activity or usage is met by users. The value of this threshold should not be too small as it would favor shorter trip patterns, and not too big as it would exclude too many users [Schneider *et al.,* 2013]. In this respect, this study used similar filtering rules, as follows:

- A day is valid for a user if he/she has a CDR in at least eight of the 48 half-hour time slots in one day (24 hours).
- Weekdays and weekends are treated separately since we presume that trip behavior can vary between them.

After filtering the 1,279,291 users extracted from the home location estimation, there remain 797,329 users that have at least one valid user-day observation (62%). This translates to a total of 4,385,089 valid user days (3,252,971 user weekdays and 1,132,118 user weekends), which correspond to 14,744,180 trips for user weekdays, and 4,965,739 trips for user weekends. This results to an average of 4.5 trips per user per weekday, and 4.3 trips per user per weekend. The filtered sample of 797,329 users corresponds to 30% of Greater Maputo's population of 2,661,832 (from HRSL), as compared to the person-trip survey which sampled only 1.7% [JICA, 2014]. This gives the advantage of CDR data of having wider representativeness.

Table 4. Statistical summary of user sample before and after filtering valid user-days

	Before Filtering	After Filtering
Number of users	1,279,291	797,329
Number of user-days	12,059,561	4,385,089 [Weekdays: 3,252,971, Weekends: 1,132,118]
Number of trips	27,117,806	19,724,307 [Weekdays: 14,744,180, Weekends: 4,965,739]

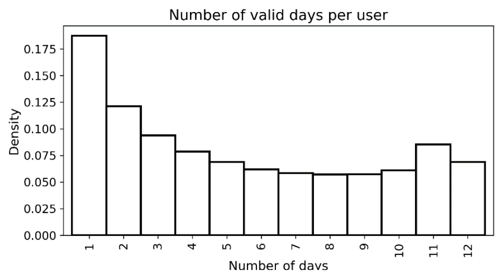


Figure 7. Number of valid days per user.

3.4 Origin-Destination Extraction

Studying macroscopic mobility require more knowledge about the start and the end of the trips to quantify trip generation and attraction volumes across different parts of a city. CDR, on the other hand, has the attribute of sparseness, which therefore require further processing in order to extract meaningful trips from. In this study, we utilize an approach similar to that of Jiang *et al.* [2017], which splits the OD extraction process into two main steps. The first is estimating the possible stay zones for each subscriber, and the second is extracting trip segments between the different stay points. The previously mentioned method is applied to triangulated CDR traces with the uncertainty of 200-300 meters for all traces [Alexander *et al.,* 2015]. However, our CDR dataset is in the cellular tower level zone which, as previously discussed, have varying coverage areas, i.e. “location uncertainty”. Spatial constraints are also added to capture more underlying trips than focusing only on key places of interest.

Extraction of Stay Locations

Based on the fact that people spend most of their time at few key locations [Isaacman *et al.,* 2011], such as their home and workplace, it is important to carefully capture those locations for each subscriber from his/her CDRs. Those locations are normally associated with a longer stay period. However, it is also important to identify other places that are visited less frequently, such as a shopping area or café, that could possibly be associated with a stay state. For that, we employed all 12 days of CDR data to extract all possible stay locations for each subscriber. Generally, for each subscriber, a zone is identified as a stay location if his/her CDRs indicate that he/she is continuously staying in a certain zone for a given time period threshold. This threshold can greatly influence the number of extracted stay locations per user, adding bias to any further trip extraction. Since cellular tower coverage varies based on population density as previously mentioned, the time period threshold should also be proportional to coverage area. The stay time period threshold impacts the number of extracted stay locations per user and correspondingly the number of trips, and should be considered reasonably. To capture trips with short stay locations, for example, buying at a store or dropping children at school, it is important to define the minimum time period threshold that would not capture a false stay location. The exact stay time varies based on multiple parameters including the used transportation mode to arrive and leave the destination and more importantly the coverage area of the subject cellular towers. Accordingly, we set the minimum time period threshold at 30 minutes in order to make sure that small stays are extracted without capturing noise. Figure 8 shows the distribution of the number of extracted stay locations per user. The resulting average number of stay locations extracted per user is 3.2.

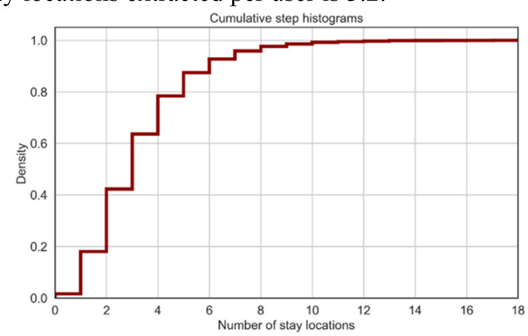


Figure 8. Distribution of number of stay locations per user

Extraction of Trips

The basic concept for trip extraction is capturing recorded locations (in terms of cellular tower zones) with successive time stamps as a trip or part of a trip depending if the locations are identified as stay locations. Figure 9 shows an example of a trip, where S1 and S2 are origin and destination stay locations, respectively, and T1 and T2 are intermediate records.

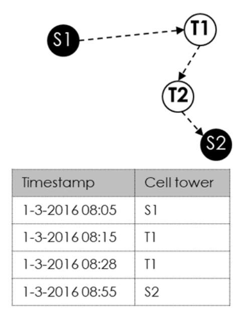


Figure 9. Example of extracted trip from CDR

3.5 Magnification Factors

In order to expand the user sample to represent the actual population of Greater Maputo, we use two types of magnification factors similarly in previous research [Jiang *et al*., 2017; Zin *et al.,* 2018]: 1) for upscaling of user sample to represent the actual population in each zone, and 2) for normalizing the user sample to one observation day.

User Sample to Population Magnification Factor

The user magnification factor for a user i in zone j (usr_mag_{ij}) is just the proportion between the total user sample with home location in that specific zone (Pop_user_j) and the actual population taken from HRSL in the same (Pop_HRSL_j), as follows:

$$usr\_mag_{ij} = \frac{Pop\_HRSL_j}{Pop\_user_j}$$

Figure 10 shows the distribution of obtained user magnification factors in the study area.

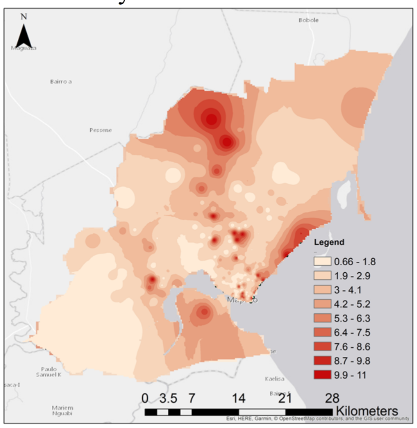


Figure 10. Distribution of user magnification factors (usr_mag_{ij}) in the study area

Valid User-Days Magnification Factor

The valid user-days magnification factor is for normalizing all users' mobility to one observation day, particularly for those with more than one valid user-day, as discussed in Section 3.1. The valid user-days magnification factor should be treated separately for weekday and weekend samples. The sample period has 9 weekdays and 3 weekends, therefore, the maximum valid user-days for the weekday sample and weekend sample are 9 and 3, respectively. Accordingly, the valid user-days magnification factor for user i ($day\_mag\_fac_i$) is just the proportion between the user magnification factor of user i (usr_mag_{ij}) and the user's corresponding valid user-days (usr_day_i), as follows:

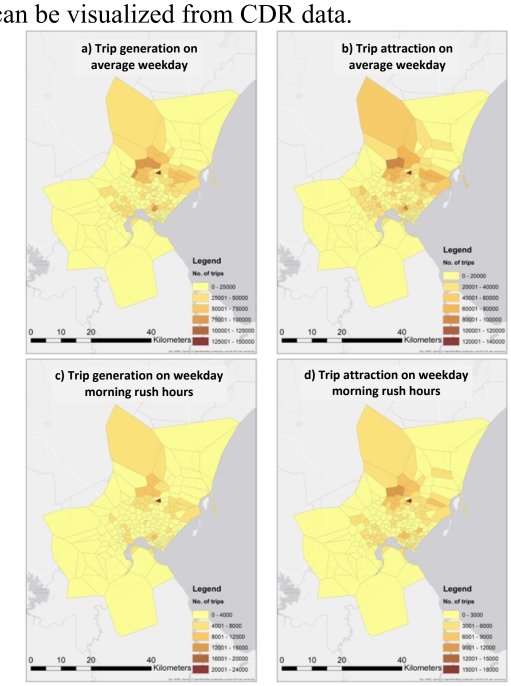
$$day\_mag\_fac_i = \frac{usr\_mag_{ij}}{usr\_day_i} \begin{cases} if\ weekday, & 1 \le usr\_day_i \le 9 \\ if\ weekend, & 1 \le usr\_day_i \le 3 \end{cases}$$

4 Results and Validation

4.1 Results

The results of our method for extracting the trips of users in Greater Maputo are presented in the form of trip generation and attraction maps, as shown in Figure 11. The trip generation and attraction maps are classified as follows: a) Trip generation on average weekday, b) Trip attraction on average weekday, c) Trip generation on weekday morning rush hours (6:00–9:00), d) Trip attraction on weekday morning rush hours (6:00–9:00), e) Trip generation on weekday evening rush hours (16:00–19:00), f) Trip attraction on weekday evening rush hours (16:00–19:00), g) Trip generation on average weekend, and h) Trip attraction on average weekend. It can be observed that all eight maps show similarity in the number of trips for all zones. The zones outward the center of Greater Maputo, particularly, in Marracuene District, Boane City, and lower part of Maputo City, have relatively low generated and attracted trips. This implies that a smaller share of the population lives in these zones, and correspondingly less travel activity. On the other hand, the central part of Maputo City and the central-northern part of Matola City shows the most number of generated and attracted trips, implying a greater share of Greater Maputo's population reside in the zones there and a high level of travel activity. This is in fact the case as these zones are part of the city center, wherein the central business district, main commercial areas, and Mozambique's major university are located. These zones continually produce and attract both short and long trips. These maps also provide an insight on the land use of these zones, such as for residential, business, or education.

Figure 12 shows the resulting people flow maps separately for a) weekday and b) weekend, which show the origin and destination of trips connected by lines, and Figure 13 shows the resulting OD matrix, having 259 origins/destinations. It can be observed in the flow map that most of the trips are heavily concentrated in the central part of the study area, similar to the observation from the trip generation and attraction maps, and that there is an obvious difference in trip volume between weekdays and weekends, i.e., the latter being much lower. Also, there seems to be a formation of four clusters of short trips in the central area, which suggests land use with heavy daily activities in both weekdays and weekends. In addition, it can also be observed that some of the trips cross water, which are presumably the trips made by water ferries. From the report of JICA [2014], 1% of the total trips are made using this transport mode. It is interesting to note that this aspect can be visualized from CDR data.



Moreover, we can capture how much people arrive to the university at different times. It is interesting that we captured two peaks during weekdays when people arrive, i.e., in the morning (8:00–12:00) and in the evening (17:00–18:00). The first peak period pertains to the trips taken by the students (including university professors and staff) for the regular classes, while the other peak period for the evening classes mostly taken by working students, as verified from the university.

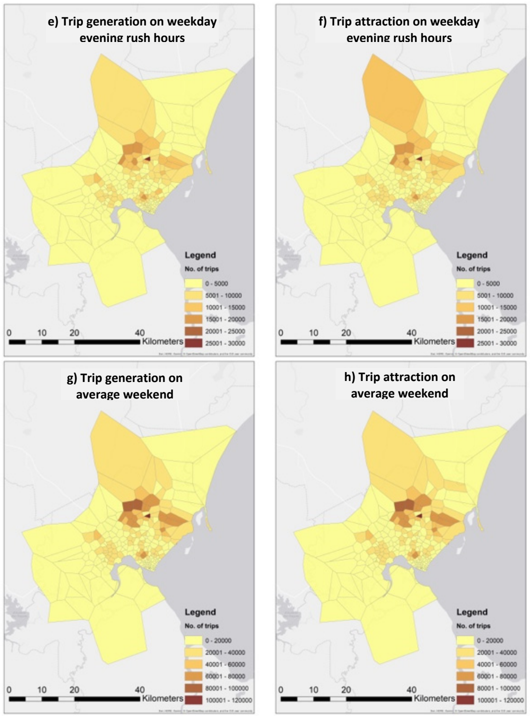


Figure 11. Trip generation and attraction maps:
a) Trip generation on average weekday, b) Trip attraction on average weekday, c) Trip generation on weekday morning rush hours, d) Trip attraction on weekday morning rush hours, e) Trip generation on weekday evening rush hours, f) Trip attraction on weekday evening rush hours, g) Trip generation on average weekend, and h) Trip attraction on average weekend

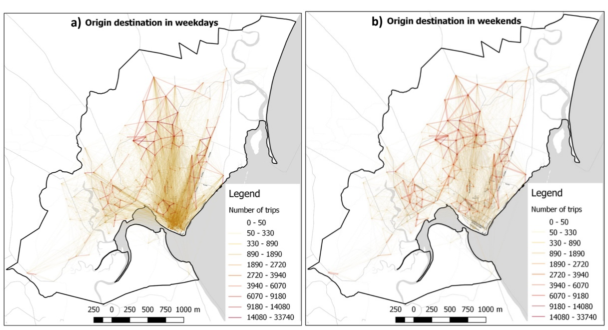


Figure 12. People flow maps in Greater Maputo:
a) weekday, and b) weekend

Further, we show a demonstrative example of using our approach in capturing accurately the origin of all trips to a specific zone as destination, and the change in mobility between weekdays and weekends. Figure 13 a) and b) shows a flow map of the trips toward Universidade Eduardo Mondlane, Mozambique's oldest and largest university, on weekdays and weekends, respectively, while Figure 14 shows the temporal distribution of trips over a 24-hour period. With this example, we can distinctly observe the difference of number of trips between weekday and weekend, i.e., there are more trips as classes are typically held on weekdays.

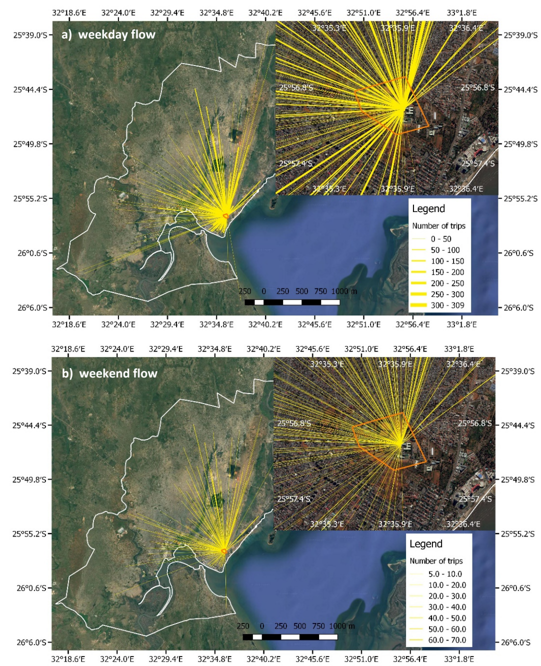


Figure 13. Flow map for trips to Universidade Eduardo Mondlane:
a) on weekday, and b) on weekend

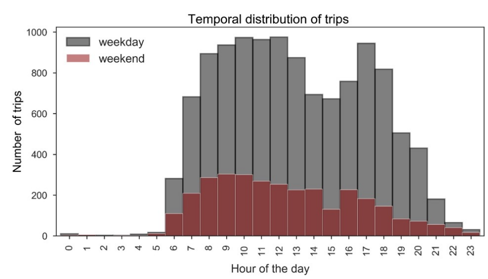


Figure 14. Temporal distribution of trips to Universidade Eduardo Mondlane

4.2 Validation of Results

For validation of results, we compare the extracted trips from CDR data with JICA's person-trip interview survey data. However, our validation has its limitations, as follows: 1) only weekday trips were considered since JICA's data did not cover weekends, 2) the JICA person-trip survey accounted for population based on the 2011 census, whereas we used population from HRSL in 2015, and 3) the Voronoi zones vary with JICA's traffic analysis zone levels (A TAZ, B TAZ, and C TAZ), as discussed in Section 3.1. Figure 15 shows the comparison between the daily weekday trip volume extracted from CDR and from JICA's person-trip survey. Relatively good correlation of the daily trip volume from CDR with the B TAZ level trips (R = 0.84) and A TAZ level trips (R = 0.97) can be obtained. It can be observed that the correlation improves as the zoning size increases, considering that the zoning mismatch decreases between the Voronoi zoning level and larger TAZ levels (B TAZ and A TAZ). Also, CDR can capture short trips between neighboring zones that the person-trip survey is not able to since it only accounts for the person's main trip of purpose, such as, typically between home and workplace. Basically the person-trip survey does not account for trips other than their main purpose. This can be seen as an advantage of CDR data over the person-trip survey data as the mobility of people taking short trips as well as the intermediate points/locations of trips can be captured, not just the "endpoints" or origin and destination locations.

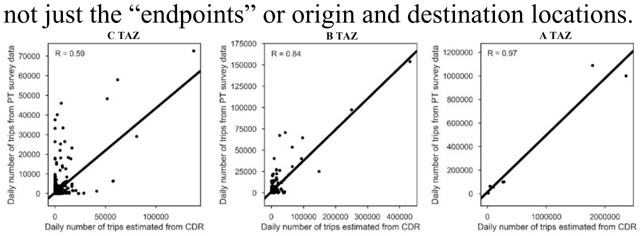


Figure 15. Comparison of daily trip volumes obtained from CDR and person-trip survey (C, B, A TAZs)

5 Conclusion

This study provided an innovative method based on proven techniques to extract spatiotemporal mobility patterns of millions of people living in a metropolitan area from big mobile phone data (CDR), and the results of which are presented into trip generation and attraction and people flow maps that may be specifically useful for urban and transportation planning purposes. Our method demonstrated the advantages of using CDR data, particularly utilizing less resources in terms of cost, time, and labor, easily implementable in a large scale, and having a larger sample, over traditional data collection methods. In addition, our method was able to capture trips in different time frames (weekday and weekend) that person-trip interview surveys are not able to, which could then lead to biased results. Our results can be practically useful for planners and policymakers by providing them some insight to which areas should be considered or prioritized for developing/improving new/existing road and public transportation networks and infrastructure. In the future, we can develop our method to obtain more accurate results, and extend the coverage of our study for the entire country of Mozambique.

References

[Alexander *et al*., 2015] Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C. González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58:240–250, September 2015.

[Cervero and Murakami, 2010] Robert Cervero, and Jin Murakami. Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and planning A*, *42*(2):400-418, February 2010.

[Facebook and CIESIN, 2016] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe. Accessed 4 January 2018.

[Isaacman *et al*., 2011] Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, Margaret Martonosi, James Rowland, and Alexander Varshavsky. Identifying important places in people's lives from cellular network data. In *Proceedings of the 9th International Conference on Pervasive Computing*, pages 133-151, San Francisco, USA, June 2011. Springer, Berlin, Heidelberg.

[Jiang *et al*., 2017] Shan Jiang, Joseph Ferreira, and Marta C. Gonzalez. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data*, 3(2):208–219, June 2017.

[JICA, 2014] Japan International Cooperation Agency (JICA). Comprehensive Urban Transport Master Plan for the Greater Maputo: Final Report, Executive Summary. March 2014.

[McNally, 2007] Michael G. McNally. The four-step model. In *Handbook of Transport Modelling: 2nd Edition*, pages 35–53, September 2007. Emerald Group Publishing Limited.

[Schneider *et al*., 2013] Christian M. Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84), July 2013.

[Phithakkitnukoon *et al*., 2010] Santi Phithakkitnukoon, Teerayut Horanont, Giusy Di Lorenzo, Ryosuke Shibasaki, and Carlo Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Proceedings of First International Workshop on Human Behavior Understanding*, pages 14–25, Istanbul, Turkey, August 2010. Springer, Berlin, Heidelberg.

[Wang *et al*., 2010] Huayong Wang, Francesco Calabrese, Giusy Di Lorenzo, and Carlo Ratti. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems*, pages 318–323, Funchal, Portugal, September 2010. IEEE.

[Zegras, 2004] Christopher Zegras. Influence of land use on travel behavior in Santiago, Chile. *Transportation Research Record: Journal of the Transportation Research Board* (1898):175–182, January 2004.

[Zin *et al*., 2018] Thein Aye Zin, Ko Ko Lwin, and Yoshihide Sekimoto. Estimation of Originating-Destination Trips in Yangon by Using Big Data Source. *Journal of Disaster Research*, 13(1):6–13, February 2018.