UTTRI

Research Report

Preliminary Analysis of Montevideo Cellular Data Records

Report 4, iCity SOUTH

Ahmadreza Faghih-Imani, Eric J. Miller

January 2018

iCITY-SOUTH:
Urban Informatics for Sustainable Metropolitan Growth in Latin America

REPORT 4:
PRELIMINARY ANALYSIS OF MONTEVIDEO CELLULAR DATA RECORDS

A report to CAF, the Development Bank of Latin America.

By:

Ahmadreza Faghih-Imani, Ph.D.
NSERC Post-Doctoral Fellow, UTTRI

and

Eric J. Miller, Ph.D.
Professor, Department of Civil Engineering
Director, UTTRI

January, 2018

EXECUTIVE SUMMARY

The purpose of this report is to undertake a preliminary analysis of a one-day sample of cellular network data provided to the project by Antel, the primary cellular telecommunications company in Uruguay. These records provide time-space traces of movements of cellphone users within the Montevideo that can be used to impute travel movements by these users. The purpose of the analysis presented in this report is to provide a preliminary indication of the usefulness of cellular network data for travel analysis and modelling purposes, and to make recommendations for how a much larger, multi-day/week sample of cellular network data could be used to support transportation analysis and planning in both Montevideo and other Latin American urban regions for which similar data might be available.

The analysis undertaken that there is significant potential in cellular network data for travel modelling purposes. While the analysis in the report provided some pieces of evidence for this potential, in order to have an acceptable level of accuracy for the inferred results, more data is needed. The ideal dataset should at least have 7 days of data to cover weekdays and weekends, although data for multiple weeks would be the ideal case. Further, having the traces for every 5 minutes rather than 15 minutes will increase the accuracy of inferred activities if the data is available.

From a larger dataset, the following information is possible to be inferred with a good level of accuracy:

- Identifying meaningful locations for individuals especially the two major anchor points: home and work locations
- Generating home-work linkage matrices
- Estimating out-of-home activity locations, start times and durations
- Creating origin-destination matrices

Cellular network data provides many advantages, such as the possibility of having large datasets for long periods with almost no response burden on the users. However, there are several challenging problems with using cellular network data for travel modelling purposes. First, the spatial precision of the data is not high. The issue is worse in rural areas where the cell sizes of BTS antennas are typically large. Also, it is not possible to detect any intra-cell movements. Further, the zones' alignment of cellular network data usually does not match the traffic analysis zone boundaries and TAZs are used for most of the standard transportation planning models.

.

ACKNOWLEDGEMENTS

This study was funded by CAF. The unstinting support and patience of CAF, and, in particular, Nicolas Estupiñan and Andres Alcala has been greatly appreciated. The collaboration with Diego Hernandez, Universidad Católica del Uruguay, and Antonio Mauttone, Universidad de la República de Uruguay, has been most welcome and helpful. The generous support of Antel in providing access to the cellular data records analyzed in the report is gratefully acknowledged.

Table of Contents

LIST OF FIGURES

LIST OF MAPS

LIST OF TABLES

CHAPTER 1: STUDY PURPOSE & MOTIVATION

Urban regions within Latin America (and elsewhere) face enormous challenges in terms of the provision of transportation infrastructure and services to meet the travel needs of their growing population in a cost-effective, equitable and sustainable manner. High quality, comprehensive information concerning travel behaviour and transportation system performance is a fundamental prerequisite for successful urban transportation planning and decision-making to address these pressing, first-order needs.

In recognition of this need, CAF established the Urban Mobility Observatory (OMU, *Observatorio de Movilidad Urbana*)[1] to assemble and utilize standardized transportation-related data for Latin American cities. 29 cities are currently members of OMU. Collecting consistent, time-series data for these cities, however, is a difficult and costly task for CAF and its partner cities.

At the same time, exciting, new transportation data collection sources are emerging to complement or even replace the traditional methods used to collect the OMU data. These include:

- The pervasive penetration of cellphone and smartphone technology within urban populations.
- The widespread adoption of smartcard systems by public transit agencies in many cities.
- Extensive deployment of many types of sensors (video, thermal, Bluetooth, etc.) for monitoring travel flows.
- Increasing availability of very large (typically crowd-sourced) datasets collected in a variety of ways by private sector companies (Google, Waze, Inrix, etc.) that can provide travel information.
- Web-based survey methods to complement/replace traditional survey methods such as home-interviews, telephone interviews, etc.

In 2015, the University of Toronto Transportation Research Institute (UTTRI) launched the *iCity* research program, which is dedicated to applying modern *urban informatics* (the combination of data collection, data science, modelling, visualization and high-performance computing methods) to the promotion of sustainable metropolitan growth. As one component of CAF's strategy for promoting its urban sustainable mobility objectives, it has partnered with UTTRI to create the *iCity-South* research program to apply the *iCity* urban informatics vision and capabilities in Latin American cities.

Two initial projects were chosen to launch the *iCity-South* research program. One involves the demonstration of agent-based microsimulation methods for modelling urban travel demand in terms of developing a prototype microsimulation model for Asunción, Paraguay.[2] The second is investigating traditional and new data collection methods in Montevideo, Uruguay. This report is the fourth in a series of reports documenting the Montevideo project results.

This report presents a preliminary analysis of a one-day sample of cellular network data provided to the project by Antel, the primary cellular telecommunications company in Uruguay. The cellular

[1] https://www.caf.com/es/temas/o/observatorio-de-movilidad-urbana/

[2] This project was completed in April, 2017. See Miller, et al., (2017a, 2017b) for the results of this project.

network records several cell to infrastructure connection events namely *handovers* (HO), *location updates* (LU) and *call detail records* (CDR). HO and CDR provide data of communication events such as calls or SMSs and LU notify the cellular network when a cellphone moves from one BTS to another. These records provide time-space traces of movements of cellphone users within the Montevideo that can be used to impute travel movements by these users. The purpose of the analysis presented in this report is to provide a preliminary indication of the usefulness of cellular network data for travel analysis and modelling purposes, and to make recommendations for how a much larger, multi-day/week sample of cellular network data could be used to support transportation analysis and planning in both Montevideo and other Latin American urban regions for which similar data might be available.

In addition to this brief introduction, this report consists of four chapters that are organized as follows. Chapter 2 briefly reviews the literature and summarizes the advantage and disadvantage of using cellular network data for travel analysis. Chapter 3 describes the cellular network data employed in this report. Chapter 4 presents the results of the preliminary analysis on the sample of cellular network data. Finally, the fifth chapter concludes the report with some recommendations.

CHAPTER 2: LITERATURE REVIEW

With the advance of technology, investigating people movements using a device which they carry has become popular in recent years due to the potential to obtain accurate data for a large sample with relatively low cost (Lee *et al.*, 2016). One approach is to use GPS services on people's smartphones or providing a dedicated GPS device to individuals to carry around during the data collection period. Another technology that makes it possible to track people movements is cellular network data from the base transceiver stations (BTSs). The increasing market penetration of cellphones provides a very large sample of data with widespread coverage, especially for urban regions.

One benefit of cellular network data is the ability to use existing infrastructure. It is not required to install new antennas or devices in contrast to other technologies such as Bluetooth. Unlike the methods using GPS services, for using cellular network data, there is no need to ask people to install an app on their phones. All travel by all cellphone users (not just smartphone owners) can be potentially tracked using the information recorded from BTS antennas. This benefit also highlights the privacy concerns associated with this type of data. It is not possible for users to opt-out or to refuse to participate in the data collection as long as they use their cellphones. Another issue with this type of data is that individual characteristics such as socio-demographic attributes are unknown. Therefore, if the population of cell phone users are not representative of the entire population, it is not possible to correct for the sample bias. Also, if a person carries two cellphones, there is no way to identify that person and not count the movements twice.

Each BTS antennas serves a certain area, usually referred to as a BTS cell. The spatial precision of the cellular network data depends on the size of this cell. While in dense urban areas the cell size is small, the cell size in rural areas can reach several kilometers. Also, in rural areas, the service coverage is not usually complete and will result in gaps in the cellular network data. This lack of spatial precision increases the level of uncertainty and makes the cellular network data more suitable for aggregate traffic measures such as generating OD matrices, estimating road congestion or investigating long-distance travel.

Table 1 Benefits and Disadvantages of CellularNnetwork Data

Advantages	Disadvantages
Use of existing data	No individual information
Large sample size	Depends on cell phone market penetration
Large coverage	Multiple counting of a person with multiple devices
Long time period	Lack of spatial precision
No interaction with users	

Earlier studies have examined the use of cellular network data for transportation planning purposes from various dimensions. Several studies have reviewed current methods and practices, potentials and limitations of using cellular network data for transportation planning analyses (Caceres *et al.*, 2008; Jiang *et al.*, 2013; Çolak *et al.*, 2015; Wang *et al.*, 2017). Studies investigating cellular network data to better understand human mobility patterns mostly focus on identifying activities, trips, and spatial-temporal variations in travel patterns (Becker *et al.*, 2011; Bekhor and Shem-

Tov, 2015; Pucci *et al.*, 2015; Xu *et al.*, 2017; Zahedi and Shafahi, 2017). Specifically, detecting home locations is considered important to yield useful insights into people's travel patterns. Although passive cell phone location data can be sparse in both space and time, inferred home locations are found to be relatively accurate as they are locations which are repetitively visited by people. The identified home locations further are used as anchors to examine other activities and trips (Ahas *et al.*, 2010; Xu *et al.*, 2015).

Another classical use of cellular network data in literature is estimating origin-destination (OD) matrices. However, since the BTS cells are not typically aligned with traffic analysis zones or census tracts and must be re-aggregated or redistributed, the level of accuracy decreases for transportation analysis. Nevertheless, it is possible to obtain BTS cell-to-cell OD matrices from cellular network data. In the transportation literature, several studies have focused on extracting OD matrices from cellular network data for different regions all around the world (Caceres *et al.*, 2007, 2013; Zhang *et al.*, 2010; Calabrese *et al.*, 2011; Mellegard *et al.*, 2011; Iqbal *et al.*, 2014; Nanni *et al.*, 2014; Larijani *et al.*, 2015; Pucci *et al.*, 2015; Demissie *et al.*, 2016). The extracted OD matrices are then used for different purposes, such as optimizing public transport network service (Berlingerio *et al.*, 2013) or estimating traffic flows traffic flows (Gundlegård *et al.*, 2016). Several recent studies have employed cellular network data to develop activity-based travel demand models (Alexander *et al.*, 2015; Pozdnoukhov *et al.*, 2016; Yin *et al.*, 2017).

In the field of traffic engineering, cellular network data are also used to estimate traffic flows and to derive routes traveled. In urban areas where the BTS cell sizes are relatively small, it is possible to assign movements to the road network. Earlier studies have worked on estimating volume and speed, improving traffic assignments and inferring real-time traffic information using cellular network data (Caceres *et al.*, 2012; Tettamanti and Varga, 2014; Janecek *et al.*, 2015; Toole *et al.*, 2015; Wu *et al.*, 2015). Due to uncertainty caused by the larger BTS cell size and the possibility of having several roads within a cell for the movements' direction, most of these studies looked at aggregated measures of traffic. Further, cellular network data has also been used for route choice problems as it can provide middle points along the way between origins and destinations.

Given the level of spatial precision, several recent studies found that cellular network data is best suited for evaluating long-distance travel patterns and obtaining large samples of data for those trips (Bekhor *et al.*, 2013; Janzen and Axhausen, 2015). Although in recent years the number of long-distance trips is growing due to better access, these trips are usually under-reported in traditional travel surveys, which typically collect only one day of people's travel diaries. Cellular network data allows passive data collection for a long period and thus can better represent less frequent trips such as long-distance travels.

In order to evaluate the potentials of cellular network data for travel analysis and modelling purposes, this report investigates the cellphone data of small sample of population in Montevideo in three main dimensions based on the above mentioned earlier studies: 1) identifying home locations, 2) inferring stops and activities, and 3) generating trips and developing OD matrices.

CHAPTER 3: DATA

Data used for the analysis in this report consists of the location of 5,000 cellphone users provided to the project by Antel, the primary cellular telecommunications company in Montevideo, Uruguay. The locations are recorded every 15 minutes for one day from 4:00 am for 24 hours. The dataset has 480,000 records, 96 time-space traces for 5,000 cellphone users. About 5.4% of the dataset includes records which the cellphone is off or out-of-zone. These records are mostly during night period and the most available data is around noon. Figure 1 shows the distribution of missing data over the 24 hours of the cellular network data sample.

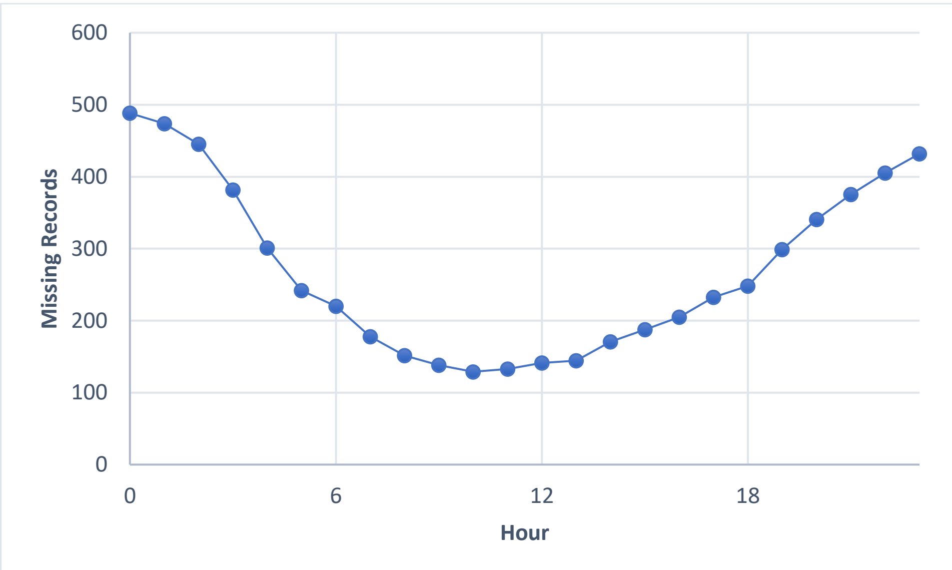


Figure 1 Distribution of Missing Records

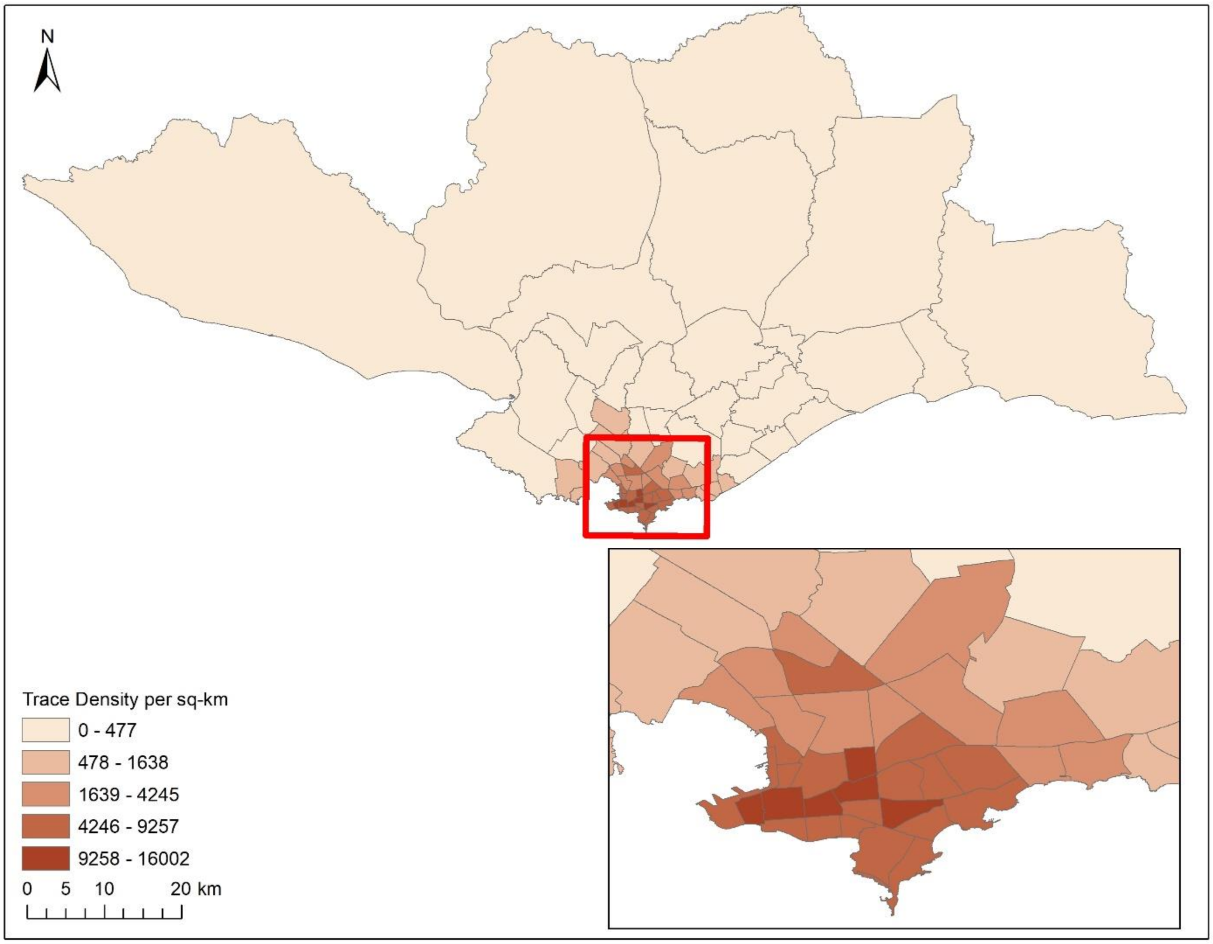
The locations in the dataset are aggregated to a zonal level. The region is divided into 74 zones. Table 2 presents the number of traces recorded for each zone for the day of the sample along with zone areas. As expected, zones 2 and 5 that represent the Central Business District (CBD) in Montevideo have the most and zone 205 that is the rural area of San Antonio and San Bautista has the least recorded cell phone traces per area.

Table 2 Data Summary

ID	Zone ID	Number of Records in dataset	Area (km)	ID	Zone ID	Number of Records in dataset	Area (km)
1	2	8458	0.608	**38**	200	1152	26.306
2	5	11428	0.714	**39**	201	2551	1100.790
3	6	5434	0.626	**40**	202	5235	59.846
4	7	5325	0.720	**41**	203	2355	782.503
5	8	3926	0.500	**42**	204	600	422.543
6	15	3775	0.670	**43**	205	114	380.338
7	16	3736	0.404	**44**	206	6887	17.914
8	30	4973	2.259	**45**	207	561	52.472
9	31	4747	1.973	**46**	208	236	514.149
10	39	10849	1.054	**47**	209	3853	15.924

11	43	6976	1.408	**48**	210	3032	25.745
12	44	7159	1.199	**49**	211	12178	3.273
13	48	6259	1.810	**50**	212	7879	1.729
14	53	5887	3.594	**51**	213	6405	1.931
15	114	2413	3.270	**52**	214	9471	6.122
16	158	934	187.151	**53**	215	13431	2.697
17	162	1668	96.473	**54**	216	12155	2.088
18	170	1305	17.054	**55**	217	6907	1.376
19	173	918	20.641	**56**	218	15538	1.168
20	174	1048	16.771	**57**	219	7770	1.008
21	183	1763	136.311	**58**	220	14226	2.034
22	184	1715	60.123	**59**	221	13493	2.479
23	185	762	584.027	**60**	222	5643	2.323
24	186	6703	1.384	**61**	223	3315	2.765
25	187	1532	52.428	**62**	224	4765	4.100
26	188	3168	30.054	**63**	225	7688	12.375
27	189	3609	0.695	**64**	226	9384	4.077
28	190	11358	15.648	**65**	227	4943	6.681
29	191	4043	71.086	**66**	228	7278	1.715
30	192	8667	8.890	**67**	229	13385	2.340
31	193	9754	0.847	**68**	230	9708	4.773
32	194	2958	91.661	**69**	231	8959	8.930
33	195	5973	17.462	**70**	232	17062	8.803
34	196	8543	11.337	**71**	233	9212	31.416
35	197	5546	19.748	**72**	234	4419	9.257
36	198	8753	15.105	**73**	235	5242	13.023
37	199	4834	2.326	**74**	236	10221	0.833

Map 1 shows the zoning system. The smallest zone in the study area is 0.4 km^2 and the largest zone is about 1,100 km^2. Map 1 also illustrates recorded traces per area for each zone for the entire one day of the sample. The patterns are expected as the denser areas have more cellphone trace records per area than rural areas.



Map 1 Zones and distribution of traces across region

CHAPTER 4: INVESTIGATING POTENTIAL OF CELL PHONE DATA FOR TRANSPORTATION MODELLING

This chapter investigates the potentials of cellular network data for transportation planning analyses with a focus on the three main objectives described in Chapter 2; i.e. identifying home locations, inferring activities, and based on the first two items generating trips and OD matrices. This section presents the processing and methodology applied to the cell phone data for these analyses.

The cellular network data include the traces of cellphone users for every 15 minutes. Following a person for the entire day, it is possible to track movements and identify stationary times. This chapter presents the main results of this report and is organized as follows: First, the analysis on detecting home locations as the most important anchor point for individuals' travels is presented; second, the stops and activities done in a day are located; and finally, the trips are generated and OD matrices are created. The limitations of each procedure as well as recommendations for future works are discussed at the end of each section.

4.1 HOME LOCATIONS

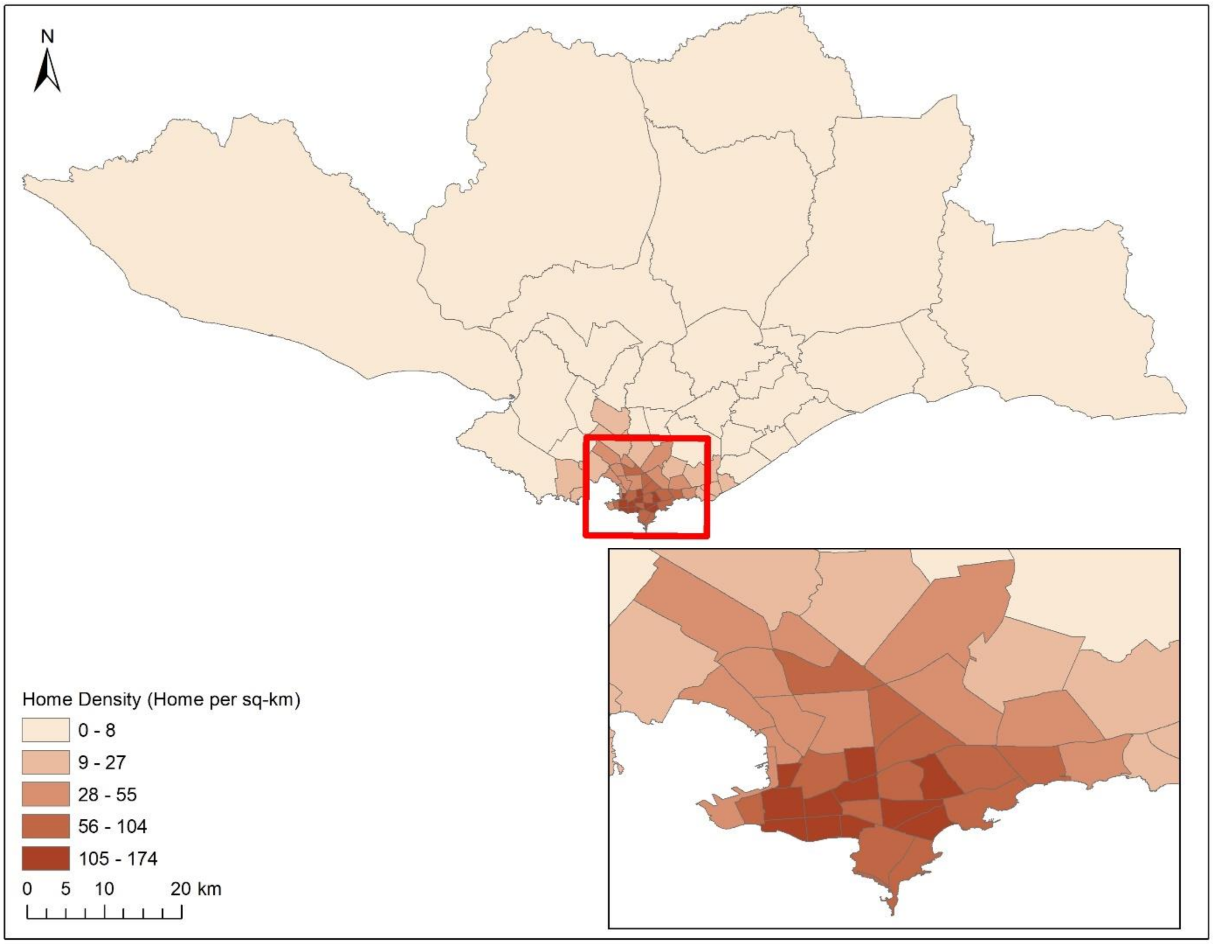
Home location is of special importance for transportation planning as most travel starts from or ends at home, while other activities are also planned around the home location. Unlike traditional travel surveys, wherein home and work locations are typically acquired in the survey, cellular network data does not explicitly have any individual-level information. However, cell phone location traces allow us to infer and make high accuracy guesses on people's primary locations such as their place of residential location.

For example, Table 3 presents the cellphone location traces of one individual in a day. This person starts the day (4:00 am) at zone 221 and stays there until 7:45 am. Then, in the evening around 7:00 pm, the individual returns to the zone 221 and remains there for the rest of the day and overnight. Based on this pattern, it is very likely that the cellphone user's home is located in zone 221.

Table 3 Example of a cellphone user's traces during a day in dataset – identifying home

Time	4:15	4:30	…	7:45	8:00	…	18:45	19:00	…	24:00	…	3:45
Zone	221	221	…	221	230	travelling different zones	230	221	…	221	…	221
	Home				Other locations and activities			Home				

With multiple days of data and observing a similar pattern of starting the day at a certain zone and going back to the same zone at night, it is safe to identify that zone within which a person's home is located. Based on this rationale, an algorithm is developed to identify zones that the cellphone users are located most of the time during the night period, i.e. 1:00-5:00 am. By looking at that period in the data sample provided to us, the location traces for two consecutive nights are being considered since the dataset starts from 4:00 am and ends the next day at 3:45 am. The procedure results in identifying home locations for 4,803 cell phone users. For the remaining 197 cellphone users, 24 hours of data are not enough to identify a zone where the user mostly records traces during the night period. Map 2 illustrates the number of detected homes per area for each zone in the study area. It is important to note that only the 4,803 identified home locations are used for this map.



Map 2 Identified Home Location Density

Sensing other meaningful locations to cellphone users, especially work/school locations, would be possible with multiple days of traces, which would allow regularly visited places to be detected. Also, having multiple days of cellular network data would increase the certainty for the identified home locations, especially for those who reside near the zones' boundaries. For those users, the overnight traces might consist of two nearby zones, as the cellphone occasionally connects to two nearby antennae. Multiday data of traces will help to distinguish between the two zones and more confidently locate the home zone. In the literature, one study has investigated 12 months of data for more than 0.5 million cellphone users and developed and tested algorithms to identify main anchor points such as home and work locations (Ahas *et al.*, 2010). Another study used those algorithms for 13 days of cellphone location data for more than 1 million users and inferred home locations (Xu *et al.*, 2015). We recommend at least 7 days of data with the date information available to be able to estimate home and work locations with an acceptable level of certainty.

4.2 STOPS AND ACTIVITIES

The next step after identifying major anchor points, such as home location, is to understand cellphone users' activities. By using a long dataset with many days of traces, locations that are visited multiple times can be identified. This helps to differentiate between recorded traces when travelling or pursuing an activity. For example, Table 4 presents the rest of traces for the person in the example in Section 4.1. The user leaves the home at zone 221 around 8:00 am and travels until reaching a destination (probably work location) at zone 218. The user remains at this location until noon when he/she goes to another location, zone 212, for one hour and then returns to zone 218 again to continue the previous activity or start a new one. The user stays at activity 3 until 18:15 pm and then arrives home around 19:00 pm.

Table 4 Example of a cellphone user's traces during a day in dataset – identifying activities

Time	…	7:45	8:00	8:15	8:30	…	12:00	12:15	…	13:15	13:30	…	18:15	18:30	18:45	19:00	…
Zone	…	221	230	229	218	…	218	212	…	212	218	…	218	236	230	221	…
	Home		Travel		Activity 1			Activity 2			Activity 3			Travel		Home	

In this report, due to the limited size of sample, the information to confidently identify any location or activity other than home location is not available. However, to demonstrate the usefulness of cellular network data to detect travel patterns, it is assumed that any location that a user stays more than 30 minutes is because of making an activity. With this assumption, for the 5,000 cellphone users, a total of 32,790 activities or an average of 6.56 activities per day is estimated. Figure 2 shows the distribution of the inferred activities over the 24 hours of the sample data. The activity trend is expected with the maximum activities happening during PM peak period.

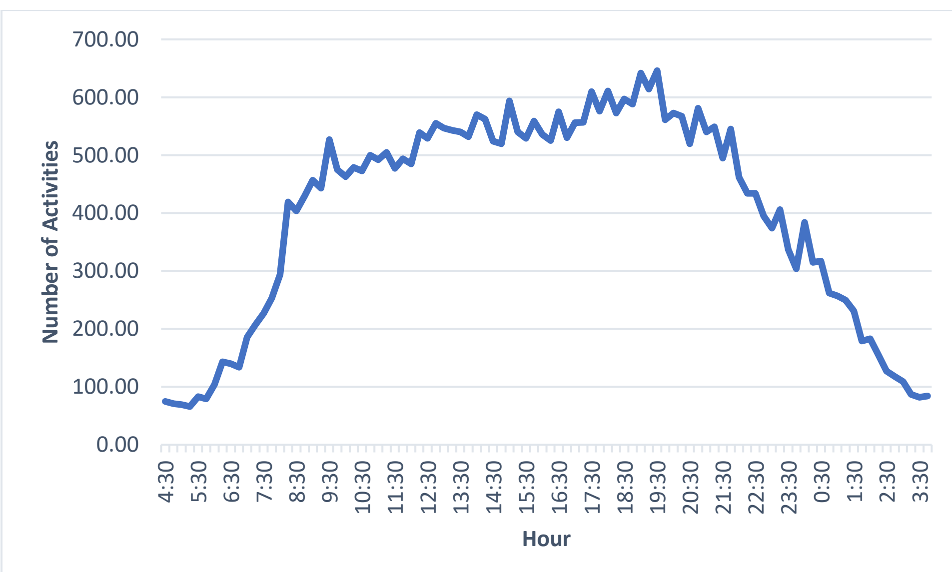


Figure 2 Inferred Activities in a Day

It is important to note that with one day of data, it is not possible to make any high-accuracy guess of the activity types. With several days of data, first, the anchor locations which a user visits multiple times can be identified and then the activities can be generated more accurately. Further, a more temporally detailed dataset with records every 5 minutes would be preferable in order to have a better realization of activity start and end times as well as a better understanding of shorter activities. Nevertheless, the efforts in this section demonstrate that by mining information from cellular network data, it is possible to infer peoples' activities.

4.3 TRIPS

In order to pursue an out-of-home activity, an individual must travel to the activity location. The previous section has identified people's activities from the cellular network data. Based on these detected activities, trips are generated between each activity. One activity location is the origin of the trip to another activity destination. For example, the user in the example in Table 4 has made four trips in that day: Home to activity 1, activity 1 to activity 2, activity 2 to activity 3, and activity 3 to home.

Each generated trip has its own start and end time and an origin and destination based on the detected activities which are connected by the trip. Thus, it is possible to create Origin-Destination (OD) matrices for each time period based on the generated trips. It is important to note that the accuracy of OD matrices depends on the accuracy of activities estimated and again one day of data is not enough to reach an acceptable level of accuracy. However, the exercise in this report aims to only demonstrate the potentials of cellular network data for travel modelling purposes.

The number of trips generated for each cellphone user is equal to the number of activities generated for that person plus one. This results in 37,790 trips for the 5,000 individuals in the dataset, i.e. an average of 7.56 trips per day. The trip rate is more than the typical trip rates from the common travel data collection methods. This inflated rate is due to errors in activity identification. For example, often and especially when the activity location is near boundaries of BTS cell, the cellphone switches from one antenna to another antenna without the user actually moving. Another example is when our naïve algorithm identifies a long trip in a large zone as an activity because it took more than 30 minutes to travel in that zone. Such errors will be minimized by using multiple days of data as the accuracy of detecting point of interest locations such as home and work places increases. In the literature, the generated trip rates from multiple days of cellular network data are very close to the rates obtained from the conventional travel data collection methods (Alexander *et al.*, 2015; Yin *et al.*, 2017). Figure 3 shows the number of generated trips originated and destined at every hour for the day of the sample.

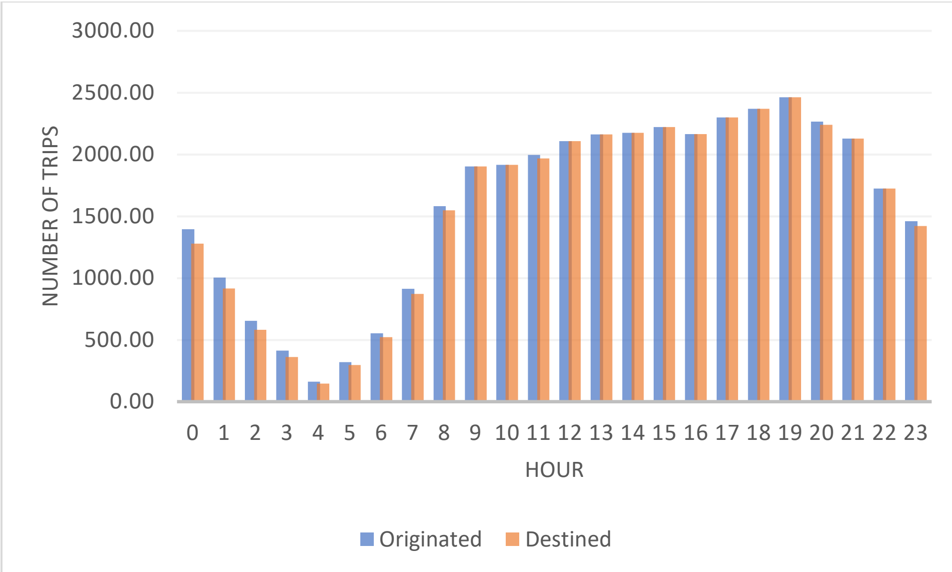
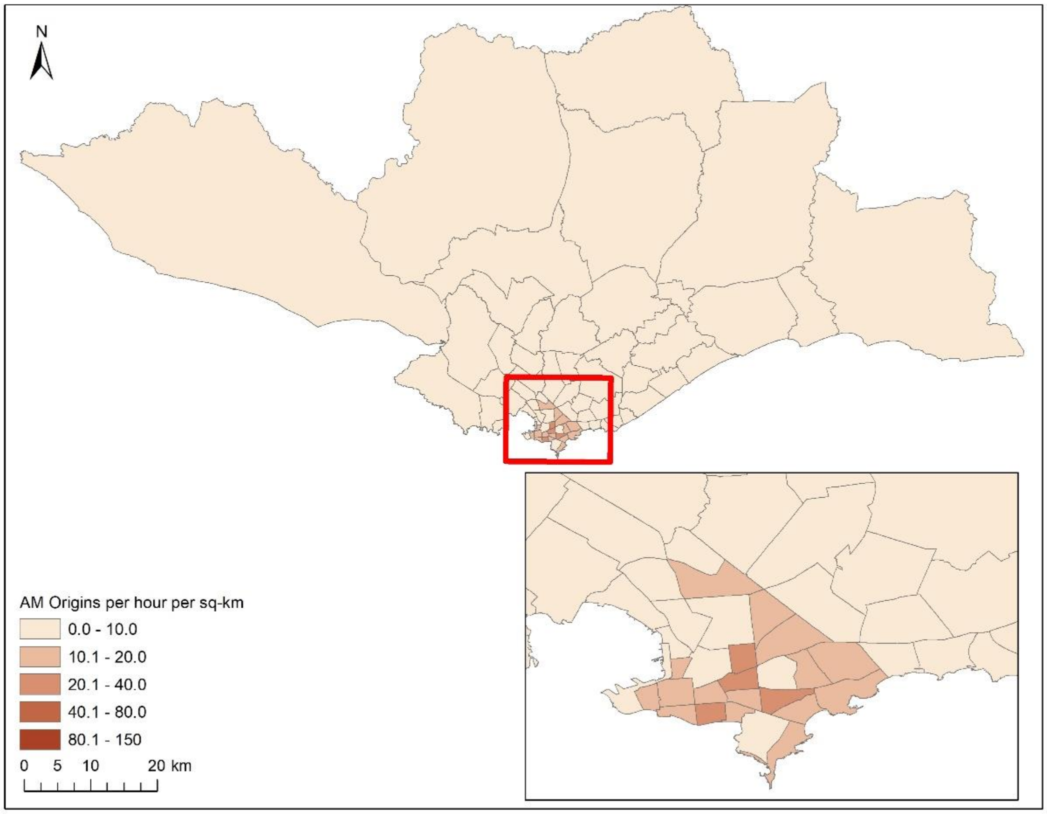
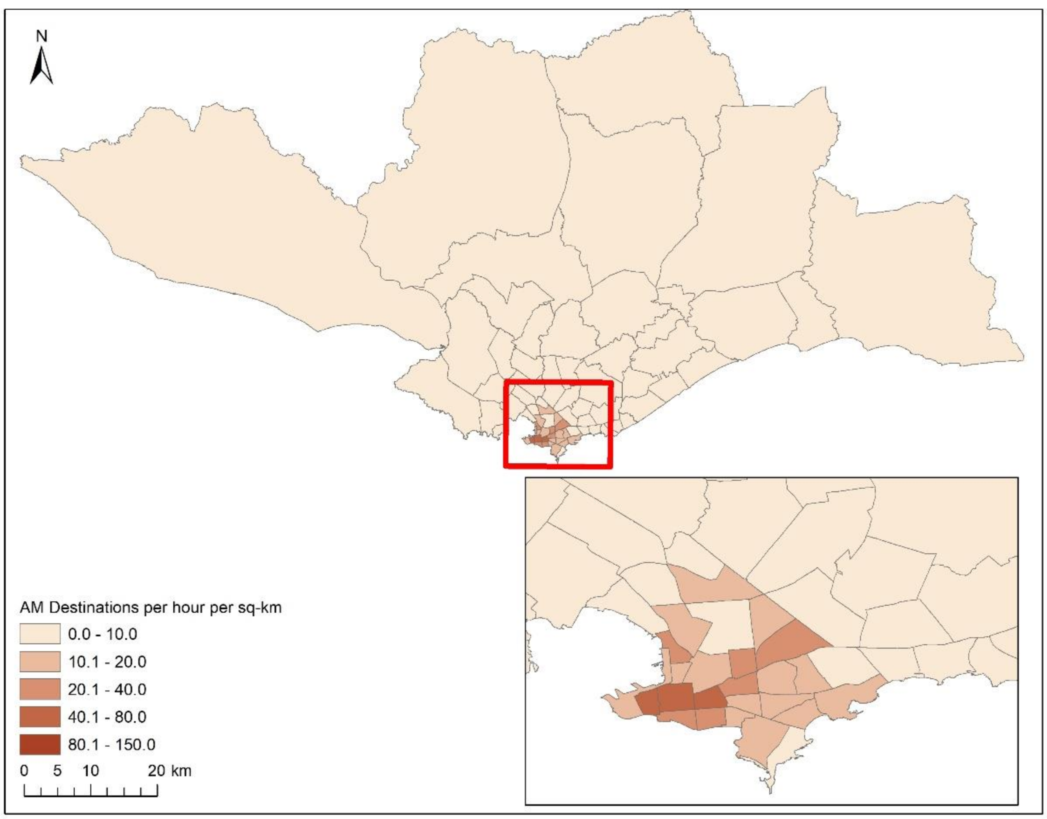


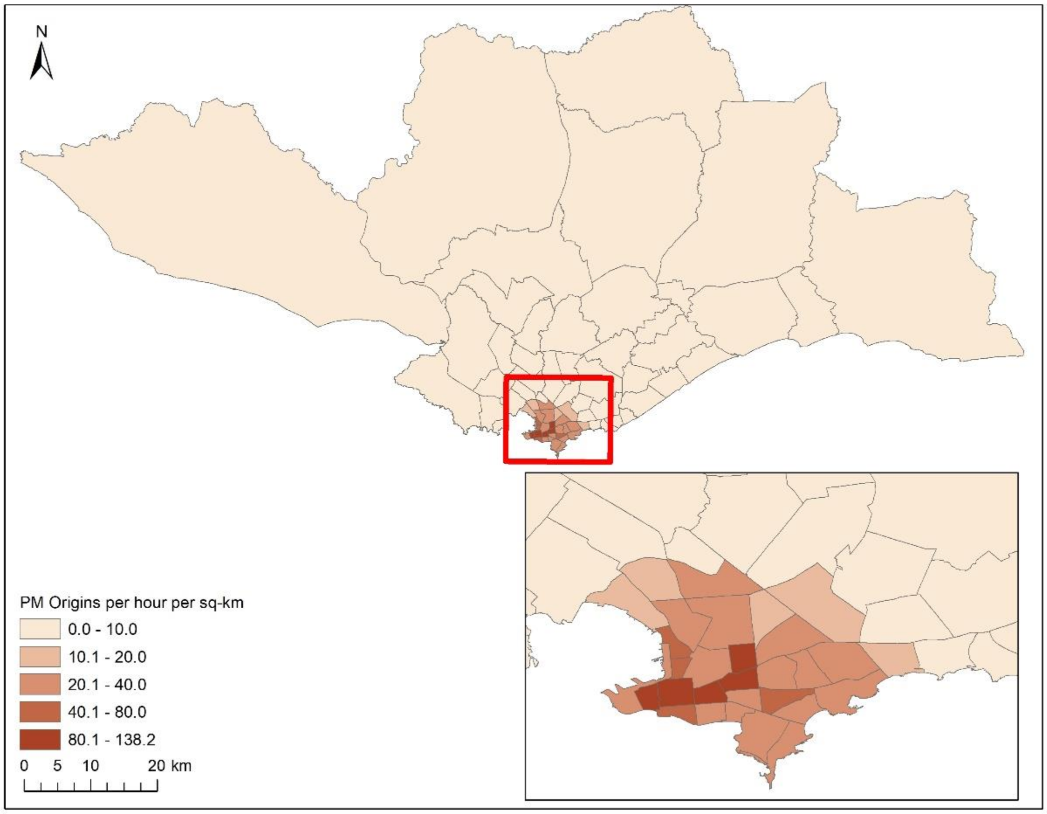
Figure 3 Distribution of Trips in a Day

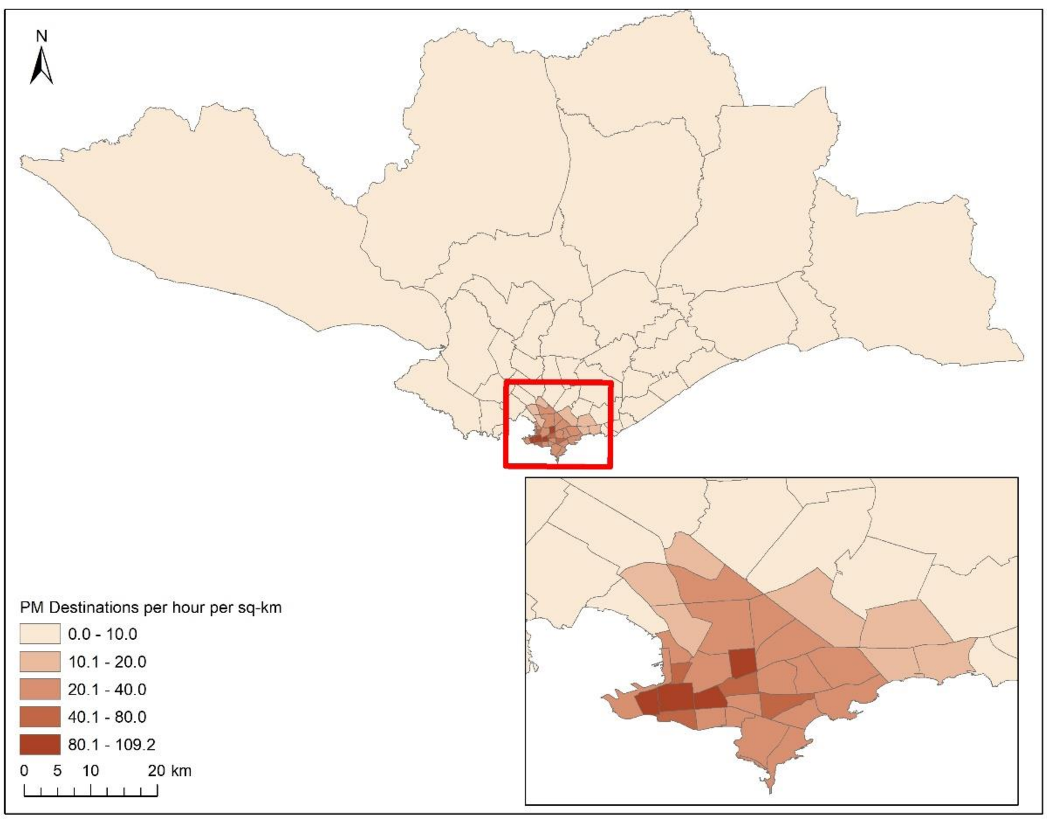
Map 3 and 4 illustrate the trips' origins and destinations aggregated to two peak periods. The AM peak is defined from 7:00 to 9:00 am and PM peak is defined 16:00 to 19:00 PM. The maps indicate an overall higher rate of trips in the PM period. In the AM period, the concentration of trip destinations in the CBD area is higher than the trip origins while in the PM the opposite is true but not as significantly as AM. This is expected as the cellphone traces should provide patterns similar to the people's actual travel patterns. It is important to mention that the temporal resolution of the OD matrices completely depends on the temporal details of the recorded traces. In the sample data provided for this report, it is possible to create OD matrices for 15 minutes intervals. For generating maps in Map 3, the trips are aggregated to an hourly level.





Map 3 Origins and Destinations in AM





Map 4 Origins and Destinations in PM

CHAPTER 5: CONCLUSIONS

This report presents a preliminary analysis of cellphone users' locations recorded at every 15 minutes for 5,000 individuals in Montevideo. The analysis shows that there is significant potential in cellular network data for travel modelling purposes. While the analysis in the report provided some pieces of evidence for this potential, in order to have an acceptable level of accuracy for the inferred results, more data is needed. The ideal dataset should at least have 7 days of data to cover weekdays and weekends, although data for multiple weeks would be the ideal case. Further, having the traces for every 5 minutes rather than 15 minutes will increase the accuracy of inferred activities if the data is available. In addition, observations for more than 5,000 cellphone users would be very useful. The larger the sample size, the more useful the data are to provide statistically reliable estimates of travel behaviour in the region.

From a larger dataset, the following information is possible to be inferred with a good level of accuracy:

- Identifying meaningful locations for individuals especially the two major anchor points: home and work locations.
- Generating home-work linkage matrices.
- Estimating out-of-home activity locations, start times and durations.
- Creating origin-destination matrices.

Cellular network data provides many advantages, such as the possibility of having large datasets for long periods with almost no burden on the users. Again, the larger this dataset, both in terms of the number persons included and the length of time over which they are observed, the better.

However, there are several challenging problems with using cellular network data for travel modelling purposes. First, the spatial precision of the data is not high. The issue is worse in rural areas where the cell sizes of BTS antennas are typically large. Also, it is not possible to detect any intra-cell movements. Further, the zones' alignment of cellular network data usually does not match the traffic analysis zone boundaries and TAZs are used for most of the standard transportation planning models.

REFERENCES

Ahas, R., Silm, S., Järv, O., Saluveer, E. and Tiru, M. (2010) 'Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones', *Journal of Urban TechnologyOnline) Journal*, ISSN homep, pp. 1063–732. doi: 10.1080/10630731003597306.

Alexander, L., Jiang, S., Murga, M. and González, M. C. (2015) 'Origin–destination trips by purpose and time of day inferred from mobile phone data', *Transportation Research Part C: Emerging Technologies*. Pergamon, 58, pp. 240–250. doi: 10.1016/J.TRC.2015.02.018.

Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A. and Volinsky, C. (2011) 'A Tale of One City: Using Cellular Network Data for Urban Planning', *IEEE Pervasive Computing*, 10(4), pp. 18–26. doi: 10.1109/MPRV.2011.44.

Bekhor, S., Cohen, Y. and Solomon, C. (2013) 'Evaluating long-distance travel patterns in Israel by tracking cellular phone positions', *Journal of Advanced Transportation*, 47(4), pp. 435–446. doi: 10.1002/atr.170.

Bekhor, S. and Shem-Tov, I. B. (2015) 'Investigation of travel patterns using passive cellular phone data', *Journal of Location Based Services*. Taylor & Francis, 9(2), pp. 93–112. doi: 10.1080/17489725.2015.1066515.

Berlingerio, M., Calabrese, F., Di Lorenzo, G., Nair, R., Pinelli, F. and Sbodio, M. L. (2013) 'AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data', in. Springer, Berlin, Heidelberg, pp. 663–666. doi: 10.1007/978-3-642-40994-3_50.

Caceres, N., Romero, L. M. and Benitez, F. G. (2013) 'Inferring origin-destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data', *Journal of Advanced Transportation*, 47(7), pp. 650–666. doi: 10.1002/atr.187.

Caceres, N., Romero, L. M., Benitez, F. G. and del Castillo, J. M. (2012) 'Traffic Flow Estimation Models Using Cellular Phone Data', *IEEE Transactions on Intelligent Transportation Systems*, 13(3), pp. 1430–1441. doi: 10.1109/TITS.2012.2189006.

Caceres, N., Wideberg, J. P. and Benitez, F. G. (2007) 'Deriving origin–destination data from a mobile phone network', *IET Intelligent Transport Systems*, 1(1), p. 15. doi: 10.1049/iet-its:20060020.

Caceres, N., Wideberg, J. P. and Benitez, F. G. (2008) 'Review of traffic data estimations extracted from cellular networks', *IET Intelligent Transport Systems*, 2(3), p. 179. doi: 10.1049/iet-its:20080003.

Calabrese, F., Di Lorenzo, G., Liu, L. and Ratti, C. (2011) 'Estimating Origin-Destination Flows Using Mobile Phone Location Data', *IEEE Pervasive Computing*, 10(4), pp. 36–44. doi: 10.1109/MPRV.2011.41.

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R. and González, M. C. (2015) 'Analyzing Cell Phone Location Data for Urban Travel', *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies , 2526, pp. 126–135. doi: 10.3141/2526-14.

Demissie, M. G., Antunes, F., Bento, C., Phithakkitnukoon, S. and Sukhvibul, T. (2016) 'Inferring origin-destination flows using mobile phone data: A case study of Senegal', in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, pp. 1–6. doi: 10.1109/ECTICon.2016.7561328.

Gundlegård, D., Rydergren, C., Breyer, N. and Rajna, B. (2016) 'Travel demand estimation and network assignment based on cellular network data', *Computer Communications*. Elsevier, 95, pp. 29–42. doi: 10.1016/J.COMCOM.2016.04.015.

Iqbal, M. S., Choudhury, C. F., Wang, P. and González, M. C. (2014) ‘Development of origin–destination matrices using mobile phone call data’, *Transportation Research Part C: Emerging Technologies*. Pergamon, 40, pp. 63–74. doi: 10.1016/J.TRC.2014.01.002.

Janecek, A., Valerio, D., Hummel, K. A., Ricciato, F. and Hlavacs, H. (2015) ‘The Cellular Network as a Sensor: From Mobile Phone Data to Real-Time Road Traffic Monitoring’, *IEEE Transactions on Intelligent Transportation Systems*, 16(5), pp. 2551–2572. doi: 10.1109/TITS.2015.2413215.

Janzen, M. and Axhausen, K. W. (2015) ‘Long-Term-C-TAP Simulation: Generating Long Distance Travel Demand for a full Year’. Available at: https://trid.trb.org/view.aspx?id=1337766 (Accessed: 30 November 2017).

Jiang, S., Fiore, G. A., Yang, Y., Ferreira, J., Frazzoli, E. and González, M. C. (2013) ‘A review of urban computing for mobile phone traces’, in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp ’13*. New York, New York, USA: ACM Press, p. 1. doi: 10.1145/2505821.2505828.

Larijani, A. N., Olteanu-Raimond, A.-M., Perret, J., Brédif, M. and Ziemlicki, C. (2015) ‘Investigating the Mobile Phone Data to Estimate the Origin Destination Flow and Analysis; Case Study: Paris Region’, *Transportation Research Procedia*. Elsevier, 6, pp. 64–78. doi: 10.1016/J.TRPRO.2015.03.006.

Lee, R. J., Sener, I. N. and Mullins, J. A. (2016) ‘An evaluation of emerging data collection technologies for travel demand modeling: from research to practice’, *Transportation Letters*. Taylor & Francis, 8(4), pp. 181–193. doi: 10.1080/19427867.2015.1106787.

Mellegard, E., Moritz, S. and Zahoor, M. (2011) ‘Origin/Destination-estimation Using Cellular Network Data’, in *2011 IEEE 11th International Conference on Data Mining Workshops*. IEEE, pp. 891–896. doi: 10.1109/ICDMW.2011.132.

Nanni, M., Trasarti, R., Furletti, B., Gabrielli, L., Mede, P. Van Der, Bruijn, J. De, Romph, E. De and Bruil, G. (2014) ‘Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast’, in. Springer, Cham, pp. 15–25. doi: 10.1007/978-3-319-04178-0_2.

Pozdnoukhov, A., Campbell, A., Feygin, S., Yin, M. and Mohanty, S. (2016) ‘San Francisco Bay Area: The SmartBay Project - Connected Mobility’, in *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, pp. 485–490. doi: 10.5334/baw.83.

Pucci, P., Manfredini, F. and Tagliolato, P. (2015) ‘Daily Mobility Practices Through Mobile Phone Data: An Application in Lombardy Region’, in. Springer, Cham, pp. 27–70. doi: 10.1007/978-3-319-14833-5_3.

Tettamanti, T. and Varga, I. (2014) ‘Mobile Phone Location Area Based Traffic Flow Estimation in Urban Road Traffic’, *Columbia International Publishing Advances in Civil and Environmental Engineering*, 1(1), pp. 1–15. Available at: https://pdfs.semanticscholar.org/8556/6877ef5621033aba68474a127cfa85c4f85f.pdf (Accessed: 1 December 2017).

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A. and González, M. C. (2015) ‘The path most traveled: Travel demand estimation using big data resources’, *Transportation Research Part C: Emerging Technologies*. Pergamon, 58, pp. 162–177. doi: 10.1016/J.TRC.2015.04.022.

Wang, Z., He, S. Y. and Leung, Y. (2017) ‘Applying mobile phone data to travel behaviour research: A literature review’, *Travel Behaviour and Society*. Elsevier. doi: 10.1016/J.TBS.2017.02.005.

Wu, C., Thai, J., Yadlowsky, S., Pozdnoukhov, A. and Bayen, A. (2015) ‘Cellpath: Fusion of cellular and traffic sensor data for route flow estimation via convex optimization’, *Transportation Research Part C: Emerging Technologies*. Pergamon, 59, pp. 111–128. doi: 10.1016/J.TRC.2015.05.004.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z. and Li, Q. (2015) ‘Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach’, *Transportation*. Springer US, 42(4), pp. 625–646. doi: 10.1007/s11116-015-9597-y.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z. and Li, Q. (2017) ‘Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data’. doi: 10.1080/00045608.2015.1120147.

Yin, M., Sheehan, M., Feygin, S., Paiement, J.-F. and Pozdnoukhov, A. (2017) ‘A Generative Model of Urban Activities from Cellular Data’, *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15. doi: 10.1109/TITS.2017.2695438.

Zahedi, S. and Shafahi, Y. (2017) ‘Estimating activity patterns using spatio-temporal data of cell phone networks’, *International Journal of Urban Sciences*. Routledge, pp. 1–18. doi: 10.1080/12265934.2017.1331139.

Zhang, Y., Qin, X., Dong, S. and Ran, B. (2010) ‘Daily O-D Matrix Estimation Using Cellular Probe Data’, in. Available at: https://trid.trb.org/view.aspx?id=910539 (Accessed: 30 November 2017).