Inferring Origin-Destination Flows Using Mobile Phone Data: A Case Study of Senegal

Merkebe Getachew Demissie*, Francisco Antunes, Carlos Bento
Center for Informatics and Systems of the University of Coimbra (CISUC)
Coimbra, Portugal
demissie@dei.uc.pt; fnibau@dei.uc.pt; bento@dei.uc.pt

Santi Phithakkitnukoon*, Titipat Sukhvibul
Excellence Center in Infrastructure Technology and Transportation Engineering (ExCITE) and Department of Computer Engineering, Chiang Mai University
Chiang Mai, Thailand
santi@eng.cmu.ac.th; titipat_s@cmu.ac.th

*Corresponding authors

Abstract*—In transportation planning, estimating the movement of people between a set of origins and destinations is a challenging task. This has been inferred in two way: from data on household socioeconomic attributes in each study zone or based on the characteristics of the study zones such as population, employment, number of cars, etc. However, developing these models can be difficult, especially in the developing countries where transport planners have limited budget to collect detailed transportation data. In this study, we use mobile phone data to estimate the origin-destination (OD) flows between districts of Senegal. We have developed two approaches to estimate commuting trips, recurring travel between one´s place of residence and place of work, and irregular trips, which are not recurring, but found to be predominant in developing countries. The inferred OD flows from sample users are expanded using the total population from census data. The results demonstrate how mobile phone data can be used to sense movement of large portion of population more regularly and with reduced cost, especially, in circumstances where relevant data are unavailable or in poor supply.

***Keywords*—mobile phone data; trip generation; origin-destination; commuting; mobility; data mining**

I. INTRODUCTION

African countries are experiencing a steady trend of urbanization and a rise in population. According to UN [1], continuing population growth and urbanization are projected to add 2.5 billion people to the world's urban population by 2050, with nearly 90 percent of the increase concentrated in Asia and Africa. While urban concentrations of population can prove to be the foundation of rapid economic growth, authorities struggle to meet the service demands of urban dwellers that are most dependent on public provision of housing, education, health care, and transport facilities etc. In the Sub-Saharan countries, good transport conditions are a key factor of economic development. However, urban sprawl and economic growth will also result in accelerated motorization of cities, which, if not accompanied by the improved infrastructures and transportation planning, deteriorate service coverage and service quality Cities in the developing countries have faced difficulties in capturing existing and future movements and dynamics of their urban systems. The main reason is that most of the developing cities do not have enough budgets to collect detailed information necessary for urban and transportation planning. For example, countries in the Sub-Saharan Africa region have limited data for transport planning, where most of the cities do not collect traffic counts in a regular basis, and except for cities in South Africa, no other city in the sub-Saharan countries has household travel surveys [2]. The other reason is that most of these cities are dealing with quick changes in population and economic growth, and changes in existing urban and transportation facilities, which alter travel requirements constantly.

Currently, data from alternative sources are becoming available and can be used to detect the movement of large percentage of population more regularly, with reduced cost, and in large-scale, especially in the developing countries where relevant data are unavailable or in poor supply. Cellular networks are ubiquitous in today's cities and the high mobile phone penetration rate in the developing countries provides planners ways to opportunistically sense the presence and movement of people at fine-grained resolution [3].

In this study, we argue that despite the low development state of the transportation planning practice in the Sub-Saharan Africa countries, they may benefit from what is now about a decade of research on using mobile phone data to achieve a more efficient transport system without having to go through the expensive structure that developed countries have setup for this purpose. Within this study we use Senegal as case study country.

We explore the mobile phone data provided by the D4D Challenge [4] to dynamically infer the urban mobility patterns to improve transportation planning. Our analysis covers both commuting and irregular trips based on data collected from sample users. The inferred OD flows from sample users are expanded using expansion factors based on the population of a user´s residential district and the total population from census data.

The rest of this paper is organized as follows: Section 2 reviews related work on the use of mobile phone data for the development of mobility patterns and origin-destination flows. Section 3 gives a description of the case study area and data. In Section 4 and 5, we explain trip estimation techniques and provide discussion of results, and finally in Section 6, we present a summary of the work developed addressing the main

978-1-4673-9749-0/16/$31.00 ©2016 IEEE

conclusions, limitations and pointing out some future research directions.

II. Related work

In recent years, researchers are exploring ways to develop large scale mobility sensing by employing the increasing capabilities found in the cellular networks system [5]-[7]. It is possible to approximate the location of mobile phone users whenever they make a call or send a short message service. There are a number of studies focusing on the use of mobile phone data for travel demand estimation. Çolak et al. [8] and Alexander et al. [9] made an attempt to assign an activity type of home, work or other to each user´s stay locations in order to infer their trip purposes. This study introduces an approach that identifies types of activities, which is important for travel demand estimation. However, improved travel demand estimation requires activity types beyond home, work, and other. Toole et al. [10] applied a pool of heterogeneous data such as census, surveys, open and crowd-sourced geospatial data for travel demand estimation. A study by Alexander and Gonzlez [11] developed a new framework to evaluate the demand and congestion impacts of a new transport mode. Jiang et al. [12] applied cellphone data to develop an activity-based model in an attempt to generate an overall daily activity patterns. Demissie et al. [13] applied cellphone data to develop a methodology to estimate passenger travel demand for public transport service.

There have been a number of studies that derived origin-destination trip matrices using mobile phone data [14]-[17]. These approaches can be quite efficient due to the possibility of using mobile phone data to produce results in a faster way in contrast with data obtained through traditional methods. For example, OD estimation through surveys take a long time from the initial data collection until the computation of the final result and it only characterize one point in time [14].

Regardless of its benefits, OD estimation using mobile phone data is criticized in certain occasions: intra-area trips can not be detected in the case of a single cell tower covering large area [18]; and there are also doubts on the validity of the estimated OD matrix in circumstances when sampling rate and penetration rate are not good. One of the two ways to examine the validity of the estimated OD matrix is by checking if the estimated OD matrix fit well to a gravity model [14], [19], [20]. This method does not necessarily needs knowledge of the parameter values of the model, only a high adjusted R-squared assure its validity. Calabrese et al. [14] compared the estimated OD matrix with the available Census data from existing mobility surveys to conform its validity.

III. Data description

Senegal is used as a case study to illustrate the analysis carried out in this study. Senegal is a country located in West Africa and it covers a land area of 196,712 square kilometers. In 2013, Senegal had an estimated population of 13,508,715. Senegal is divided into 14 regions, which are further divided into 45 departments, and 123 arrondissements (districts) [21]. Our analysis in made on district level. Senegal has a high number of mobile phone users. In 2012, the number of active mobile telephone cards per 100 Senegalese inhabitants was 79% [22]. In this study, analyses were performed using mobile communication data made available by SONATEL and Orange within the D4D Challenge. In 2012, SONATEL had share market of 61% in Senegal [22]. The data are based on Call Detail Records (CDRs) of mobile phone calls and text exchanges of SONATELs customers, which correspond to the month of January 2013.

For our OD flow estimation, we used mobile phone record with granularity of district level and each record in our data includes anonymized unique caller ID, and the date and time of the call (Table I). We analyzed cellular traffic handled by 1,666 base stations. For commercial and privacy reasons, SONATEL did not provide the actual geographical coordinates of the base stations. New position for each base station is assigned uniformly in its Voronoi cell (the region consisting of all points closer to that antenna than to any other) to make it more difficult to re-identify users.

TABLE I. SAMPLE MOBILE PHONE DATA

User ID	Time stamp	Arrondissements (districts) ID
1	18-03-2013 21:30:00	5
1	18-03-2013 21:40:00	5
1	19-03-2013 20:40:00	7

IV. Trip estimation

Trip or journey is defined as a unidirectional movement from a point of origin to a point of destination [23]. Depending on the type of the origin and destination points, trips can be categorized as home-based or non-home-based. The proportion of trips by different purposes usually varies greatly with time of day.

The procedure for estimating OD matrices consists of two steps: data filtering, and trips origin and destination estimation. The original dataset contained more than 9 million unique aliased mobile phone numbers, and further data processing was done at the Orange Labs to retain users that met the following two criteria [24]: (i) users having more than 75% days with interactions per given period; and (ii) users having had an average of less than 1,000 interactions per week. These measures are taken to guarantee a high sampling rate by involving active users in our sample, and to filter out noise resulting from the extremely high number of mobile phone interactions, which are presumed to be machines, or shared phones.

A. Commuting trips

We have analyzed over 43 million mobile phone records in the month of January 2013 collected from 146352 randomly selected sample users from 123 districts of Senegal. Data from the two holidays, January 1(new year´s day) and January 24 (Prophet Birthday), are not included in our analysis.

We started our analysis by inferring commuting trips, which are usually taken between the user´s residence and work locations. To identify home location, we estimated a home-district for each mobile phone user based on the most frequent used cell tower location during the nighttime (10pm – 7am). This home location detection method was first developed by Phithakkitnukoon et al. [25] who show its reliability as they compared their result against the actual census data. To identify the work location, we first filtered out calls on Sundays in order to avoid the bias from non-working days. Then, we estimated a work-district for each user based on the most frequent used cell tower location during the daytime (8am – 7pm).

Fig.1 and Fig. 2 show the connections made between the home and work locations of sample users that resemble the inter-district (Fig.1) and intra-district (Fig. 2) commuting trips.

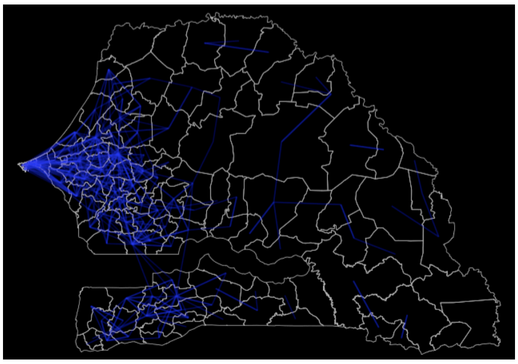

Fig. 1. Commuting trips of sample users (inter-district)

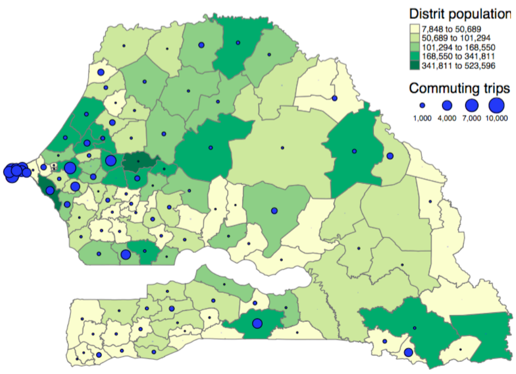

Fig. 2. Commuting trips of sample users (intra-district)

Previous studies by Nanni et al. [16] and Csáji et al. [19] also assign home and work locations for each user and inferred the commuting trips. While this approach provides ways to understand trips between home and work locations, it overlooks non-commuting trips that are available in large proportion in developing countries [23]. In the next analysis we follow an approach that accommodates all the trips and focus on the total inter-district movements.

B. Inter-district trips

For each user, we arranged the consecutive traces along the date and time of the day. Then, an inter-district trip can be recorded if mobile phone interaction of a user is detected in one district, and later another mobile phone interaction is detected at different district. Finally, trips with the same origin and destination district are grouped together at different temporal window, for example daily, weekday and weekend.

Fig. 3 shows average daily trips of sample users for the days of the week. We summed the number of trips from all the districts of Senegal for each day, and averages are calculated for the same days of the week. Besides Sunday, the number of average daily trips from Monday to Saturday are reasonably similar.

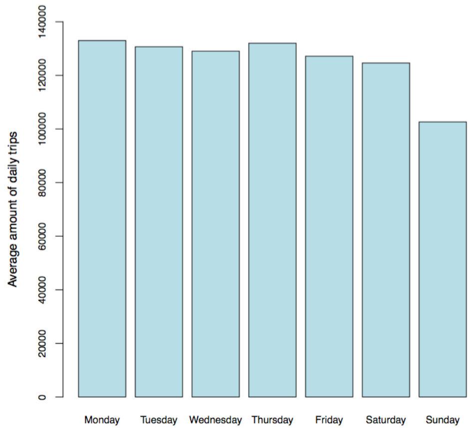

Fig. 3. Average amount of daily trips for sample users

Table II shows the distribution of average trips by sample users for all the days of the week. We summed the number of trips made by each user for the same days of the week over the month of January. Then, by dividing each user´s total number of trips by the corresponding total number of days, we obtained the distribution of average daily trips per user. The first, second and third quartile values for each average daily trips distribution were computed, revealing fewer number of trips by the users are made on Sundays.

TABLE II. AVERAGE AMOUNT OF DAILY TRIPS PER USER

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
1st Qu.	1.50	1.50	1.50	1.50	1.50	1.50	1.00
2nd Qu.	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Mean	2.58	2.61	2.57	2.61	2.56	2.58	2.40
3rd Qu.	3.00	3.00	3.00	3.00	3.00	3.00	3.00

We used the indicators from Fig. 3 and Table II in order to make decision regarding the level of aggregation to represent the inter-district trip making behavior in Senegal. There is high number of total trips as well as average number of trips per user between Mondays to Saturday compare to the corresponding number of trips on Sundays. Thus, we aggregated the trips between Monday to Saturday to represent the predominant travel patterns of Senegal, and we analyze the travel patterns on Sundays separately.

Through this process, we analyzed the average number of trips per user for two time windows, predominant travel pattern from Monday to Saturday, and weekend pattern on Sunday. Fig. 4 shows the distribution of total trips per user for 25 days between Monday to Saturday period, with first, second and third quartiles of 5, 14, and 35 trips respectively with corresponding first, second, and third quartiles of 3, 7, and 15 active trips making days. Fig. 5 shows the distribution of total trips for 4 Sundays with first, second and third quartiles of 2, 5.2, and 5 average trips respectively with corresponding first, second and third quartiles of 1, 2, and 3 active trip making Sundays.

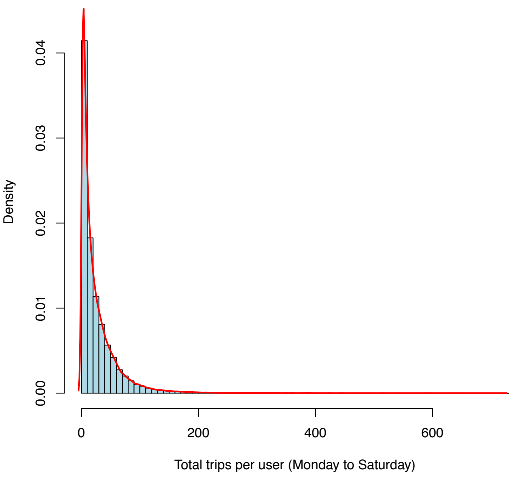

Fig. 4. Probability distribution of total trips per user (Monday to Saturday).

V. WEIGHTING AND EXPANSION OF DATA

The eventual purpose of our analysis is to draw conclusions about the mobility behavior of the population from which the sample was drawn. This can be done by means of population expansion factor, which relates the sample users to the population. In order to calculate the expansion factor, it is important to have a secondary source of data describing the population, which can also be related to the sample. In our case, we use a national Census of population for Senegal [21]. Therefore, with our mobile phone data we estimated a home-district for each user based on the most frequent used cell tower location during the nighttime (10pm-7am). Then, the expansion factor can be obtained by dividing the actual population of each district to the number of users who have been categorized as that district's residents from the mobile phone data. A similar approach is also followed in the work of Çolak et al. [8] and Alexander et al. [9] to obtain an expansion factor. However, in our approach we did not include the population of age group 0 to 4.

Fig. 6 and Fig. 9 show inter-district origin-destination flows aggregated Monday through Saturday by sample users, and the general population respectively The majority of the inter-district trips are made between districts in the region of Dakar, Thies, Saint-Louis and Diourbel, which is expected as these regions have high population density and social, economical and political significance.

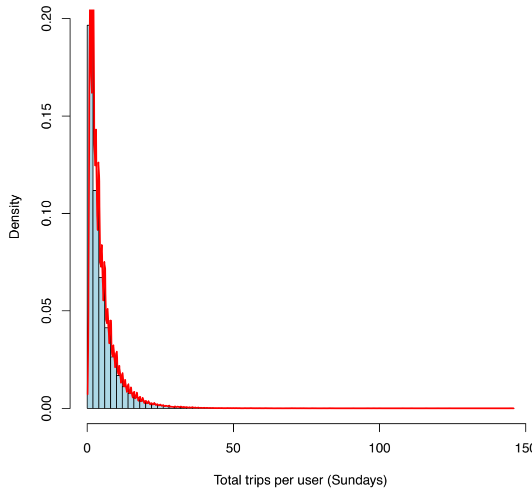

Fig. 5. Probability distribution of total Sunday trips per user.

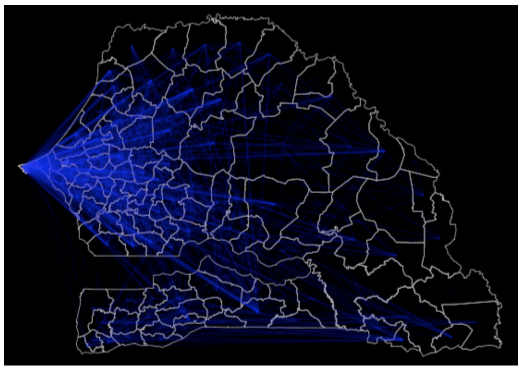

Fig. 6. Inter-district origin-destination flows by sample users (Monday to Saturday)

Fig. 7 shows how many trips originate in each district and Fig. 8 shows how many trips are destined in each district. Districts in the highly populated regions such as Dakar, Thies, Saint-Louis, Diourbel, etc. are both producing and attracting large amount of trips. On the other hand, some highly populated regions produce fewer amounts of trips (e.g. Ouro Sidy-Matam and Nyassia-Ziguinchor). Depending on availability of more data, future trip generation analysis should relate the frequency of trips to the characteristics of individuals, districts, and transportation network of each district.

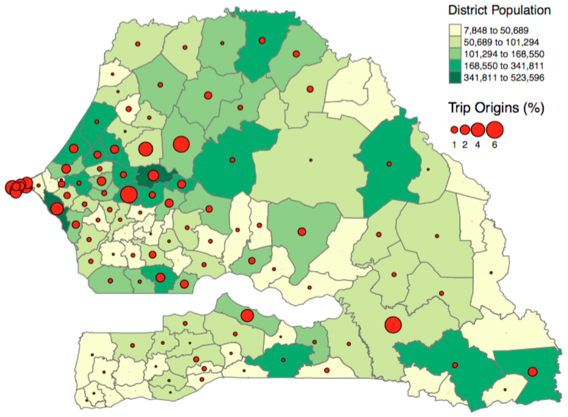

Fig. 7. Percentages of inter-district OD flows by origin expanded to the whole population (Monday to Saturday)

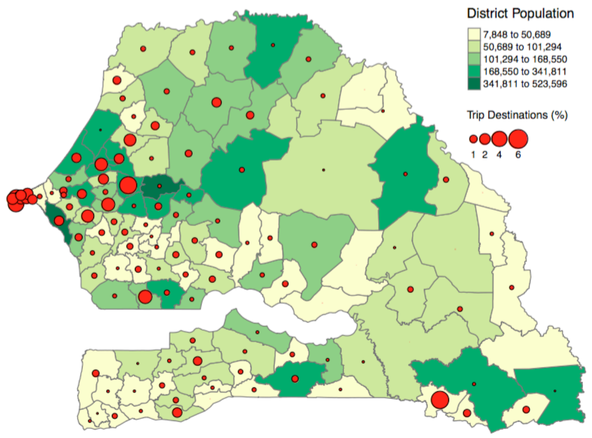

Fig. 8. Percentages of inter-district OD flows by destination expanded to the whole population (Monday to Saturday)

VI. Conclusion

Recently, given the fact that the urban share of Sub-Saharan African population has grown substantially and is expected to continue in the future, the concern for the Sub-Saharan Africa transport policy program is shifting towards improving urban mobility and accessibility [26]. However, planners in developing countries are facing great difficulties in capturing the existing and future mobility patterns due to the limited budget they can access for transport data collection purposes.

In this study, we propose to develop approaches so as to infer the urban mobility patterns by using mobile phone data obtained from Data for Development Senegal Challenge.

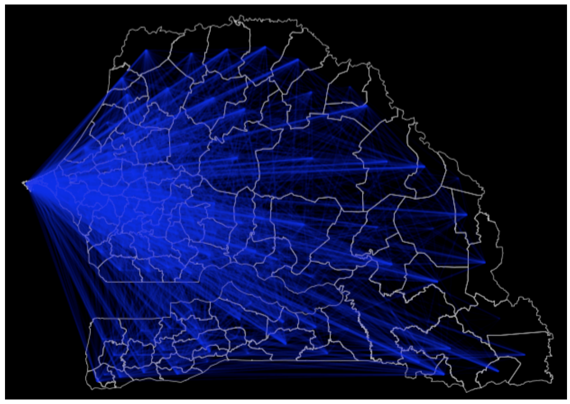

Fig. 9. Inter-districts OD flows expanded to the whole population (Monday to Saturday)

In order to demonstrate our approaches we use Senegal as case study country. We started our analysis by inferring commuting trips between 123 districts of Senegal. Commuting trips are very significant, and understanding the mobility pattern between home and work locations is very important to understanding transportation needs. However, the trip between home and work is only one of large number purposes that generate daily travel activity. Our second attempt is an analysis of inter-district flows that include both commuting and non-commuting flows. Our results enable transport planners to analyze their transport offerings. The inter-district flow information allows planners to observe the demand for an intercity public transport services.

Despite the significance of our analysis, we should highlight some important limitations of the study. In our attempt to infer the total inter-district flow is the possibility that a long tip can be divided and detected as many short trips. In order to avoid this problem previous studies by Nanni et al. [16] and Csáji et al. [19] assign home and work locations for each users and focus only on the commuting trips. While the aforementioned approach provide new ways to construct an overall daily activity patter for the activity based travel demand models, it overlooks non-commuting trips that is available in large proportion in developing countries [23], and also a similar behavior conformed by [20]. In our approach we take a risk of accommodating all the trips and focus on the total inter-district movements.

The standard approach to obtain the expansion factor require a more detailed information to explicitly account for the composition of the sample data before expanding the results to represent the population to which the sample belongs. For example, detailed information includes users socio-economic and demographic profile [27]. This approach can be used when the data in the secondary source is of a comparable level of detail to that obtained in our sample. However, for the time being, mobile phone data offers very little information about the users socio-economic and demographic profile.

Our OD estimation is based on mobile phone data, however, it may still be questioned whether the traces (CDRs) can comprehensively represent the real mobility patterns of people. This can be answered through validation with ground truth information such as mobility surveys, which is not currently available in Senegal. Future studies will have to consider integration of additional traffic information (traffic count, mobility survey, etc.) to validate the estimated OD flows.

ACKNOWLEDGMENT

The authors would like to thank the Data for Development Senegal Challenge for providing the mobile phone data. This work is supported by Suscity project financed by the FCT (MITP- TB/CS/0026/2013).

REFERENCES

[1] African development bank group, 2014. Tracking Africa´s progress in figures. Available at: http://www.afdb-org.jp/file/news-and-pressrelease/Tracking_Africafs_Progress_in_Figures-3.pdf (Accessed 01.02.2015).

[2] B. Williams, Sustainable urban transport in Africa: issues and challenges.2011, http://www.walshcarlines.com/pdf/Sustainable%20Urban%20Transport%20in%20Africa%20Brian%20Williams.pdf (Accessed 03.02.15).

[3] M.G. Demissie, "Combining datasets from multiple sources for urban and transportation planning: emphasis on cellular network data," Ph.D. dissertation, Dept. Civil. Eng., Coimbra Univ., Coimbra, Portugal, 2014.

[4] D4D Senegal challenge. http://www.d4d.orange.com/en/Accueil (Accessed 12.02.2016)

[5] M. G. Demissie, G. H. Correia, C. Bento, "Intelligent road traffic status detection system through cellular networks handover information: An exploratory study," Transportation Research Part C Emerging Technologies, vol. 32, pp.76-78, Jul. 2013. DOI: 10.1016/j.trc.2013.03.010

[6] M. G. Demissie, G. H. Correia, C. Bento, "Exploring cellular network handover information for urban mobility analysis," Journal of Transport Geography, vol. 31, pp.164-170, Jul. 2013. DOI: 10.1016/j.jtrangeo.2013.06.016

[7] M. G. Demissie, G. Correia, C. Bento, "Analysis of the pattern and intensity of urban activities through aggregate cellphone usage," Transportmetrica A Transport Science, vol. 11, no. 6, pp. 502-524, Mar. 2015. DOI: 10.1080/23249935.2015.1019591

[8] S. Çolak, L. Alexander, B. Alvim, S. Mehndiretta, M. Gonzalez. (2015). Analyzing Cell Phone Location Data for Urban Travel: Cur- rent Methods, Limitations and Opportunities. Transportation Research Board: Transit cooperative research program. Washington, DC. [Online]. Available: http://humnetlab.mit.edu/wordpress/wp-content/uploads/2010/ 10/TRB finaldraft.pdf.

[9] L. Alexander, S. Jiang, M. Murga, M. Gonzalez, "Origindestina- tion trips by purpose and time of day inferred from mobile phone data," Transportation Research Part C: Emerging Technologies, Vol. 58 (Part B), pp. 240-250, 2015.

[10] J. Toole, S. Colak, B. Sturt, L. Alexander, A. Evsukoff, M. Gonzalez, "The path most traveled: Travel demand estimation using big data resources," Transportation Research Part C: Emerging Technologies, Volume 58, Part B, Pages 161-428, 2015.

[11] L. Alexander, M. Gonzlez, Assessing the Impact of Real-time Ridesharing on Urban Traffic using Mobile Phone Data. UrbComp15, August 10, 2015, Sydney, Australia, 2015.

[12] S. Jiang, J. Ferreira, M. Gonzalez, Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. UrbComp15, August 10, 2015, Sydney, Australia, 2015.

[13] M. G. Demissie, S. Phithakkitnukoon, T. Sukhvibul, F. Antunes, R. Gomes, C. Bento. (2016). Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal. IEEE Transaction on Intelligent Transportation Systems. DOI: 10.1109/TITS.2016.2521830

[14] F. Calabrese, G. di Lorenzo, L. Liu, C. Ratti, "Estimating origin-destination flows using mobile phone location data," In: Pervasive Computing, IEEE, vol. 10, no. 6, pp.36-44, 2011.

[15] M. S. Iqbal, C. F. Choudhury, P. Wang, M. C. González, "Development of origin– destination matrices using mobile phone call data," Transportation Research Part C: Emerging Technologies, vol. 40 (2014), pp. 63–74, 2014.

[16] M. Nanni et al., "Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast." In: Citizen in Sensor Networks. Springer, pp. 15–25, 2014.

[17] J. White, I. Wells, Extracting origin destination information from mobile phone data. 11th International Conference on Road Transportation and Control, London, pp. 30-34, 2012.

[18] N. Caceres, J. P. Wideberg, F. G. Benitez, "Review of traffic data estimations extracted from cellular networks," IET Intelligent Transport Systems, 2 (3), pp. 179-192, 2008.

[19] B. Csáji, A. Browet, V. Traag, J. Delvenne, E. Huens, P. Dooren, Z. Smoreda, V. Blondel, "Exploring the mobility of mobile phone users," Physica A Statistical Mechanics and its Applications, vol. 392, no. 6, pp. 1459-1473, 2013.

[20] Y. Wang, G. Correia, E. Romph. (2014). National and Regional Road Network Optimization for Senegal Using Mobile Phone Data. Available at: http://www.dat.nl/media/uploads/files/National_and_Regional_Road_Network_Optimization_for_Senegal_Using_Mobile_Phone_Data_DAT.Mobility.pdf (Accessed 01.02.16)

[21] Geohive. (2014). [Online]. Avaialble: http://www.geohive.com/cntry/senegal ext.aspx.

[22] Deloitte and GSMA. (2012). Sub-Saharan Africa Mobile Observa- tory. [Online]. Available: http://www.gsma.com/publicpolicy/wp- content/uploads/2012/03/SSA FullReport v6.1 clean.pdf.

[23] J. D. Ortuzar, L. G. Willumsen. (1994). Modelling transport. Wiley.

[24] Y. Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, V. Blondel. (2014). D4D-Senegal: The Second Mobile Phone Data for Development Challenge.

[25] S. Phithakkitnukoon, Z. Smoreda, P. Olivier, "Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data," PLoS ONE, vol. 7, no. 6, 2012.

[26] R. Gorham., (2015). Sustainable Mobility and Accessibility in Urban Areas of Africa. Transforming Transportation 2015 conference. Washington DC, United States, January 15-16, 2015.

[27] A. J. Richardson, E. S., Ampt, A. H., Meyburg, (1995). Survey Methods for Transport Planning. Eucalyptus press. Available at:
http://libvolume3.xyz/civil/btech/semester8/urbantransportplanning/urbantransportsurvey/urbantransportsurveytutorial2.pdf (14.12.2015).