Research Article

The Potential of Mobile Network Big Data as a Tool in Colombo's Transportation and Urban Planning

Sriganesh Lokanathan, LIRNE*asia*
Gabriel Kreindler, LIRNE*asia* & MIT
Nisansa de Silva, LIRNE*asia*
Yuhei Miyauchi, LIRNE*asia* & MIT
Dedunu Dhananjaya, LIRNE*asia*
Rohan Samarajiva, LIRNE*asia*

Abstract
Transportation systems in developing economies are straining under rapid urban population growth. Better planning requires good data on population movements. A "big data"-centric approach to transportation management is already a reality in many developed economies, with transportation systems being fed with a multitude of sensor data. Developing countries by contrast rely heavily on infrequent and expensive surveys. With mobile phone usage becoming ubiquitous, even in developing countries, there is potential to leverage data from citizens' mobile phone usage for transportation planning. Such data can allow planners to produce insights quickly without waiting for the proliferation of sensors. Using mobile network big data (MNBD) from Sri Lanka, this paper explores this potential, producing mobility related insights for the capital city of Colombo. The analyses suggest population movements can be understood at a higher frequency and spatial resolution than what was possible before, complementing infrequent surveys. MNBD based insights cannot produce all the insights that are possible from specialized sensors or the infrequent surveys. But for resource-constrained economies with a dearth of timely and relevant insights, even an incremental advance in their ability to produce timely insights from MNBD can improve existing transportation planning. However more research will be required before such techniques can be mainstreamed.

1. Introduction

Transportation systems in developing economies are straining under rapid urban population growth. Road congestion is a frequent reality in many cities and in particular on the main thoroughfares that feed the city's daytime working population. Data are needed to identify the choke points and prioritize additions and enhancements. A "big data"-centric approach to transportation management, based on sensor data, is already a reality in many developed economies, with transportation systems being fed with a multitude of sensor data such as loop detectors, axel counters, parking occupancy monitors, CCTV, integrated public transport card readers, and GPS data, from phones as well as from public and private transport systems (Amini et al., 2011).

Developing countries, however, are more reliant on traditional forms of data collection such as questionnaires. Such survey-based methods, administered at peak hours, can be very costly, in terms of personnel, processing, and traffic disruption. This rather clumsy questionnaire method was used in the Sri Lankan capital Colombo in 2013. Other less intrusive methods (e.g., automatic traffic recorders) are not able to yield important information such as routes taken and parking.

Mobile network big data have enormous potential to contribute to traffic planning and complement these infrequent surveys. Because the data streams are continuously flowing, the effects of changes in traffic channels – e.g. one-way schemes and new roads – can potentially be easily tracked. And although additional costs for data storage may be involved, mobile network base transceiver station (BTS) hand-off data can even serve as trackers of traffic speed and disruptions. Moreover, as the proportion of GPS-enabled smartphones increases, it may be possible to achieve the same traffic-tracking objectives with smaller samples, i.e., without collecting masses of BTS hand-off data.

The primary question addressed by the research outlined in this article was: to what degree can mobile network big data inform transportation planning for the city of Colombo? To address this question, we attempted to use mobile network data to understand where the daytime commuting population of Colombo comes from, i.e., to create origin-destination (OD) matrices that explicate the flow of commuters between different geographic areas.

2. Literature Review: Data-Driven Transportation Management

Depending on the sophistication of the network, mobile network data can capture a range of location variables. The data can be broadly classified into two types: passive positioning data and active positioning data.

Passive positioning data are automatically generated by the mobile network and captured in the network's logs, for billing purposes and also for network management (Ahas et al., 2010). Every time a subscriber uses her or his phone to make or receive a call, to send or receive an SMS/MMS, or even to access the Internet, a mobile network BTS generates a record of that event. These records are collectively called, call detail records (CDRs). In addition to the identifiers of the parties involved in the event, date, time, and duration of the event, each record also includes the cell ID (the ID of the antenna), which in turn has a geo-location and antenna orientation information (i.e., an azimuth). These passive positioning data based on cell IDs are inexpensive when compared to active positioning data, but with the tradeoff that because they are at the level of network cells, they have a lower precision than active positioning data.

Active positioning data are location data captured via specifically initiated network queries to locate handsets, using either network-based and/or handset-based positioning methods. Location data from GPS, and/or GPS-augmented network triangulation, can also be considered active positioning data (Ahas et al., 2010). The use of these methods has come about due to national regulations (e.g., regulations requiring operators to capture high-precision location data for security reasons) and/or due to there being a business case for providing location-based services. However, not all mobile operators generate continuous active positioning data for all their subscribers, and even fewer operators store the data.

Active and passive positioning data have been utilized with great effectiveness in transportation, helping measure and model people's movements, in both developed and (to a much lesser extent) developing economies. Trip-based OD matrices (traditionally derived using infrequent surveys) have been created using mobile network big data in South Korea (Yoo et al., 2005), Spain (Caceres et al., 2007), and the United States (Calabrese, et al., 2011) amongst others. Researchers have also used Mobile network big data have also been used to conduct activity-based modeling of people's movements (Isaacman et al., 2011), to infer transportation modes (Wang et al., 2010) and to map traffic flow (Wu et al., 2013).

Even the least-developed mobile network infrastructure generates passive positioning data in the operator logs. Passive positioning cell ID data provide the least spatial resolution. Despite this, there has been much recent work using such data in transportation planning. For example, IBM researchers have utilized passive positioning CDR data from the operator Orange to map citizens' travel routes and show how data-driven insights could be used to improve planning and management of transportation services in Abidjan, the largest city in Côte d'Ivoire (Berlingerio et al., 2013). The IBM study suggests that overall travel time could be reduced by 10% through optimization of the network (in this case, by extending one bus route and adding four new ones), thus offering a partial solution to the city's congestion problems. Similar work using passive mobile positioning data to inform transportation planning and management is being conducted in other countries, both developed and developing.

3. The Data Set

Our study used four months of national passive positioning CDR data of voice calls for between 5 and 10 million SIMs from a Sri Lankan mobile operator.[1] The data were completely pseudonymised by the operator – i.e. a unique computer-generated identifier replaced each phone number – and we were not provided with any mapping information linking the phone numbers and the identifiers.

Each CDR corresponded to a particular subscriber of the operator and a CDR was created each time a subscriber originated or received a call. In the case of an in-network call (i.e., when both parties to the call were subscribers of the same mobile network), two records were generated, one for each party.

Each record contained the following data:

- call direction: a code to denote if the record was an incoming or outgoing call;
- subscriber identifier: pseudonymised identifier for the subscriber in question;
- identifier of the other party: pseudonymised identifier for the other party to the call;
- cell ID: identification of the antenna that the subscriber was connected to at the time of the call;
- date and time that the call was initiated; and
- duration of the call.

4. Data Analysis and Findings

The analyses of the data was done by assigning each SIM in the dataset a unique home and work location at the level of the divisional secretarial division (DSD), which is the third sub-national administrative level (after provinces and districts) in Sri Lanka. Based on the home and work assignments, it was possible to find regular

[1] As per our agreements with the operator, we do not name the operator or give a precise figure for the number of SIMs that were analyzed.

inter-DSD mobility patterns, which forms the basis of the analyses conducted in this paper.

4.1 Identifying Home and Work locations

Two of the motifs that characterize human mobility are “home to work” and “work to home”. Whilst there are many other motifs (e.g., “home to school”), our study utilized only these two basic motifs, because they are the most common motifs and the ones with the greatest relevance to transportation planning.

We devised a methodology to find the “home” and “work” location (in this case at the DSD-level) for each SIM. The first step in finding the home and work DSD locations was to find the time bracket within which an average individual would spend at either location. It was assumed that the time brackets falling outside the home and work brackets would be the time that individuals were spending commuting between home and work locations.

In the morning, an individual will typically leave the home location and commute toward the work location. Once the individual is at the work location, there will typically be less drastic movement. At the end of the workday, the individual will leave the work location and commute towards the home location. Once the individual is at the home location, there will typically be less drastic movement until the departure to work the next day. Thus, there should be two peaks of human movement, one in the morning when people are commuting from their homes towards their workplaces, and another in the evening when people commute from their workplaces towards their homes. The statistical valley in between the two peaks would, accordingly, be the work hours, and the times before the first peak and after the second peak would be home hours.

Based on the assumptions described above, we followed a three-step process to extract mobility graphs using data for a normal working day from the dataset:

1. Average position (latitude, longitude) for each individual SIM was calculated for each hour of the day (24 points in all).
2. The distance travelled by each individual SIM during each hour bracket was obtained by calculating the Euclidian (i.e. straight-line) distance between two consecutive hour-wise average locations.
3. An average hour-wise distance measure was then obtained for all the SIMs that showed movement during a specific hour. These results were then plotted on a graph.

Ideally this should this process should have been done for each working data in the dataset and then combined to obtain and average working day profile. But due to time limitations this was not done.

This three-step process revealed the mobility patterns illustrated below in Figure 1.

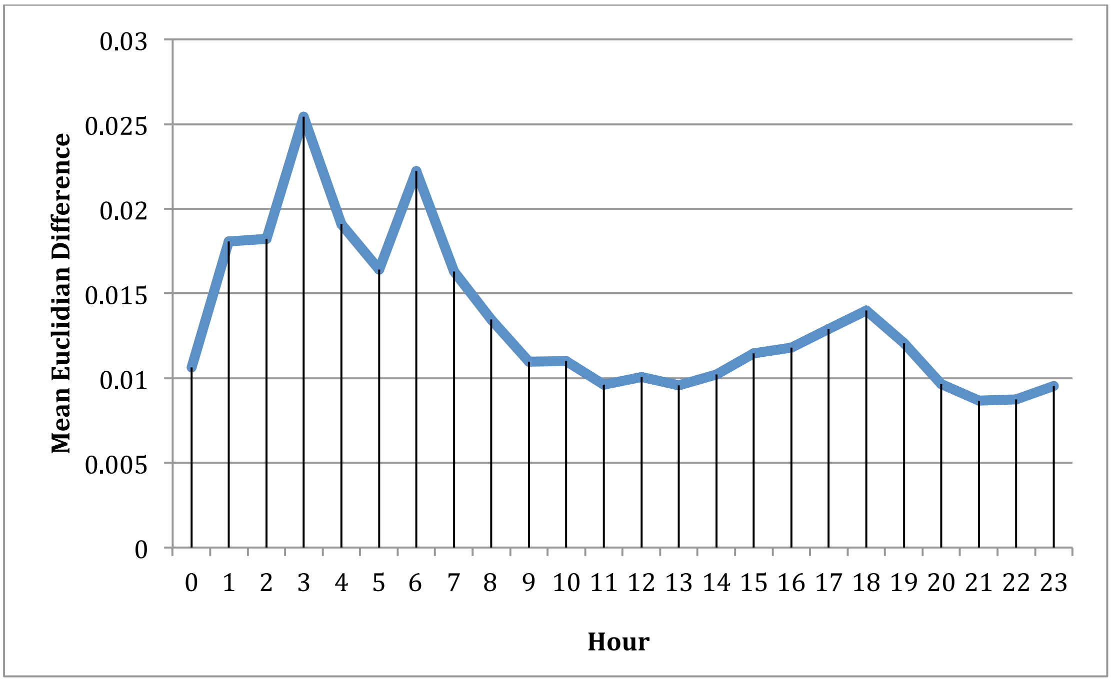


Figure 1: Average SIM mobility during hourly time brackets in a single day

Figure 1 shows, as predicted, two peaks of mobility: a morning peak at 06h00, and an evening peak at 18h00. But there is also another peak, in the middle of the night at 03h00. We determined that this spike was caused by a small number of individuals moving quickly over long distances in the middle of the night, presumably mostly truck drivers. (The reason that this minority dominated the values at that time period was that the majority of the individuals were sleeping at that time and thus not making or taking calls, resulting in null CDR values for most SIMs.)

To fix the anomaly of the 03h00 peak (and other problems arising from the sparsity of data in the middle of the night), it was decided to interpolate the position data for every individual SIM in the dataset. Thus, if an individual SIM had at least two real location values, the missing values between the real values (i.e., the null values generated by hours when there was no voice telephony activity via the SIM) were interpolated using the real values. Using the interpolated individual data, and then following the same three-step process described above, the average mobility of a SIM during each hour was calculated, as illustrated below in Figure 2.

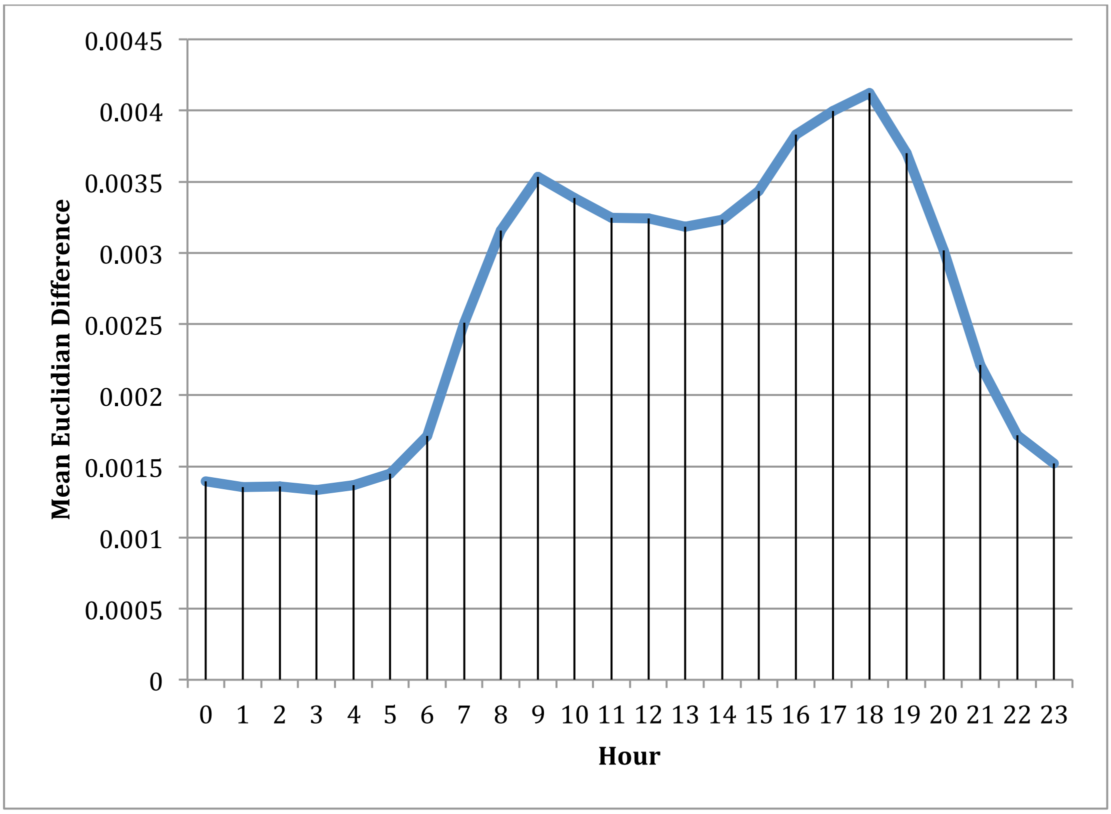


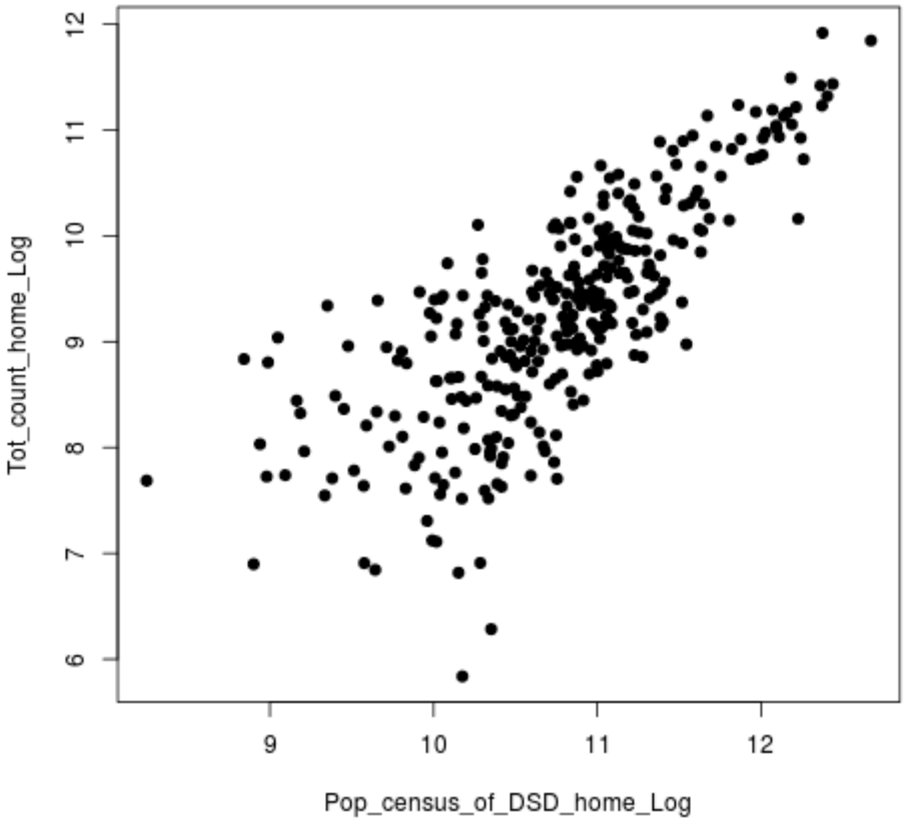
Figure 2. Average SIM mobility during hourly time brackets in a single day using interpolation

Based on this second mobility graph, it was concluded that the period 10h00 to 15h00 should be considered the work hours and the period 21h00 to 05h00 the home hours. This decision was slightly subjective, but was based on finding these to be the two most significant periods of time with minimal variation in average mobility.

Our next step was to identify home and work DSDs. To do this, we employed a variation of the "tower days" concept proposed by Isaacman, Becker, and Cáceres (2011), i.e., we used the logic behind the tower-day concept to generate a "DSD count" concept. Accordingly, assignment of home and work DSDs was carried out via a three-step process, as follows.

1. A DSD was assigned for each individual SIM for each home or work slot for each by choosing the DSD that was used the most for each of the time slots. Weekends and national holidays were left out in assigning work DSDs; for example in one 31-day month there, 10 days fell on weekends and the 2nd Tuesday was a national holiday, so for that month there were 31 home DSD allocations and 20 work DSD allocations for each SIM.
2. For each SIM, all the potential DSDs that were obtained for that SIM for the home timeslot over the 4-month time period were listed along with its associated frequency. This was also done for the SIM for the work time slot.
3. An overall Home location for each SIM was assigned to the DSD that occurred the most in the frequency table generated in the previous step for the home timeslot. The same logic was used with the frequency table associated with the work time slot to assign an overall Work location for the SIM.

In order to measure the soundness of our methodology, we compared our results with population data from the latest national census. For each DSD we compared the total number of SIMs assigned that particular DSD as a home location (a proxy for resident subscriber population in that DSD), with the DSD population from the national census. Figure 3 below shows a log-log graph where x-axis is the log of a DSD's census population and the y-axis is the log of the DSD's home location count.

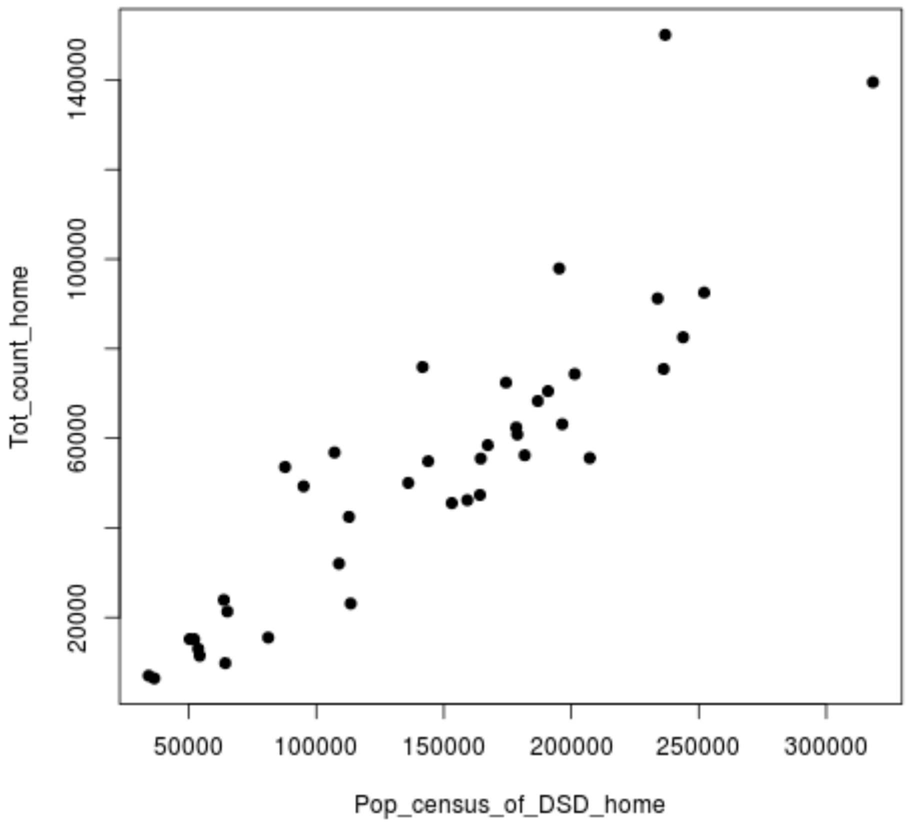


Residual standard error: 0.4664 on 325 degrees of freedom
Multiple R-squared: 0.5922
Adjusted R-squared: 0.5909

Figure 3. Correlation between derived home DSD population counts and DSD census population in Sri Lanka

The adjusted r-squared value of 0.59 whilst not low, is still not very high. This is explained by the fact that the operator doesn't have consistent subscriber penetration rates in all of the DSDs.

The mobile telephony base station density and penetration of operators were both found to be highest in the Western Province, the province that includes Colombo. Given our desire to focus this research exercise on Colombo, we decided to calculate the relationship between our home DSD counts and the census DSD data for only Western Province (Figure 4).



Residual standard error: 31630 on 38 degrees of freedom
Multiple R-squared: 0.8003
Adjusted R-squared: 0.7951

Figure 4. Correlation between derived home DSD population counts and DSD census population in the Western Province

The adjusted r-squared value of 0.79 is much better when considering only the DSDs in the Western Province rather than the DSDs in the whole country (0.59 in the latter case). This was hypothesized to be a reflection of the fact that the Western Province has a more uniform mobile penetration for the operator than the rest of the country.

4.2 Colombo City Findings

A finding from analysis of our Western Province home DSD counts was that the largest contributor to each DSD's working population came from amongst the same DSD's habitant population, i.e., that most people were living and working within the same DSD. Accordingly, in order to better understand the importance of Colombo city (comprised of the Colombo and Thimbirigasyaya DSDs) as a destination for work, we examined:

- the extent of each home DSD's contribution to Colombo's working population (i.e., each home DSDs's contribution to work DSD counts in Colombo and Thimbirigasyaya DSDs); and
- the relative rank of each home DSD's contribution to Colombo's working population.

Our analysis found that nearly 47% of Colombo city's working population came from outside the city, i.e., just over 53% of SIMs had *both* a home DSD count and a work DSD count in Colombo city (i.e., in Colombo and/or Thimbirigasyaya DSDs). Table 1 below shows the top home DSDs for the working population of Colombo city.

Table 1: Sources of Colombo city's working population

Rank	Where the working population of Colombo city lives (home DSDs)	% of Colombo's working population
1.	Colombo city (Colombo + Thimbirigasyaya DSDs)	53.07%
2.	Maharagama	3.66%
3.	Kolonnawa	3.52%
4.	Kaduwela	3.32%
5.	Sri Jayawardanapura Kotte	2.92%
6.	Dehiwala	2.60%
7.	Kesbewa	2.54%
8.	Wattala	2.47%
9.	Kelaniya	2.08%
10.	Ratmalana	1.95%
11.	Moratuwa	1.82%

When we drilled down further and looked at the two constituent DSDs that make up the city of Colombo, namely Colombo and Thimbirigasyaya, an even clearer mobility picture emerged.

Colombo DSD

For Colombo DSD, our data analysis found that nearly half (49.5%) of its working population also lived in the same DSD, with 50.5% living outside Colombo DSD (see Table 2 below).

Table 2 also shows that among the home DSDs from which Colombo DSD's working population is drawn, the Colombo DSD is either the second or third most important working location. These source DSDs are generally the DSDs north and northeast of Colombo DSD, not just from Colombo district but also from Gampaha district, which lies shares its southern boundary with Colombo district.

Table 2. Top sources for the working population of Colombo DSD

Source DSD (district)	Source (home) DSD's rank (as source of Colombo DSD workforce)	% of Colombo DSD's working population that comes from Source DSD	Colombo DSD's rank among the source DSD workforce's destinations for work
Colombo (Colombo district)	1	49.5%	1

Thimbirigasyaya (Colombo district)	2	6.0%	2
Kolonnawa (Colombo district)	3	4.0%	2
Wattala (Gampaha district)	4	3.5%	2
Maharagama (Colombo district)	5	2.8%	3
Kaduwela (Colombo district)	6	2.7%	3
Kelaniya (Gampaha district)	7	2.7%	2
Sri Jayawardanapura Kotte (Colombo district)	8	1.9%	3
Biyagama (Gampaha district)	9	1.9%	2
Kesbewa (Colombo district)	10	1.9%	3
Ja-Ela (Gampaha district)	11	1.7%	2

Also, as Table 3 below shows, the workforce of the other Colombo city DSD, Thimbirigasyaya, also gets a large contribution of its working population (47.1%) from within the same DSD, while the top outside contributors are neighbouring Colombo district DSDs (to the areas south and east of Thimbirigasyaya DSD).

Table 3. Top sources for the working population of Thimbirigasyaya DSD

Source DSD (district)	**Source (home) DSD's rank (as source of DSD Thimbirigasyaya workforce)**	**% of Thimbirigasyaya DSD's working population that comes from source DSD**	**Thimbirigasyaya DSD's rank amongst the source DSD workforce's destinations for work**
Thimbirigasyaya (Colombo district)	1	47.1%	1
Maharagama (Colombo district)	2	4.5%	2
Kaduwela (Colombo district)	3	3.9%	2
Sri Jayawardanapura Kotte (Colombo district)	4	3.8%	2
Colombo (Colombo district)	5	3.7%	2
Dehiwala (Colombo district)	6	3.4%	2
Kesbewa (Colombo district)	7	3.2%	2
Kolonnawa (Colombo district)	8	3.1%	3
Ratmalana (Colombo district)	9	2.5%	2
Moratuwa (Colombo district)	10	2.2%	2
Homagama (Colombo district)	11	2.0%	2

An interesting picture also emerges when we consider the interpolated individual data. Figure 4 below compares results from our analysis of mobile network interpolated individual data (image on left) with the results from a costly survey of conducted in the Western Province (image on right) that was used to understand mobility and transportation patterns.[2] The findings are similar.

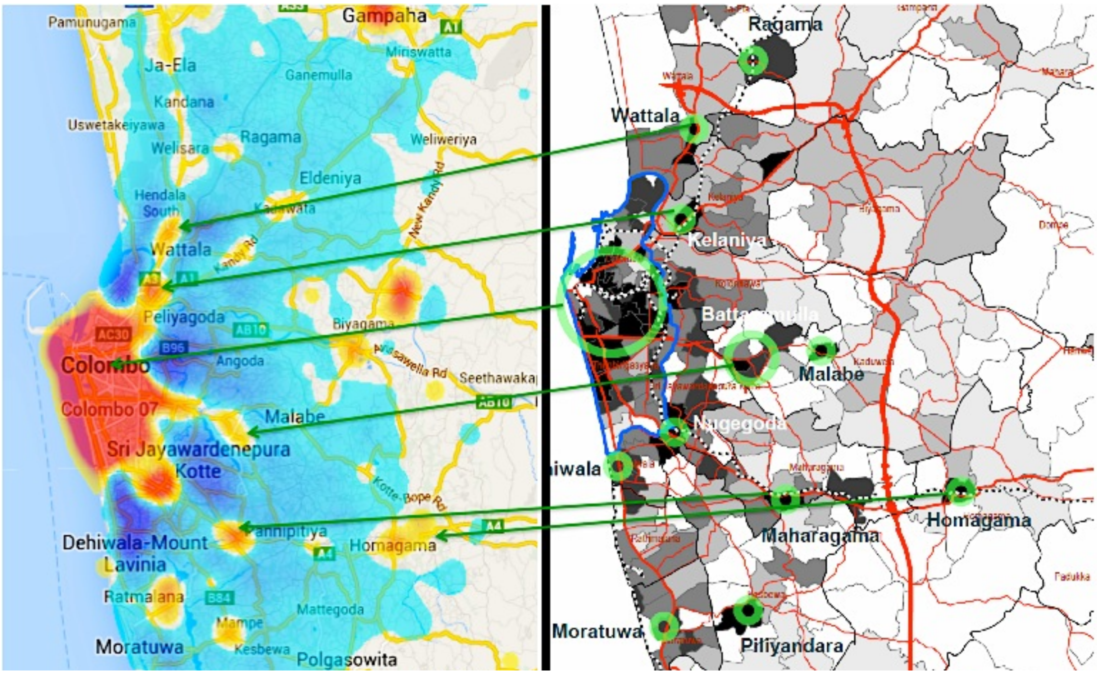


Figure 4. Results of analyses of mobile network big data (on left) and survey data (on right).

The image on the left, based on our analysis of mobile telephony data, depicts the relative population density in Colombo city and its surrounding regions at 13h00 compared to midnight (the previous day) on a weekday in 2013. Thus, the image shows the population movement from the outskirts of the city to its center during the day. The yellow to red colors show areas where density at 13h00 has increased relative to midnight (i.e., the more red, the greater the rise in density), and the blue color depicts areas where density at 13h00 has decreased relative to midnight (i.e., the darker the blue, the greater the loss in density). The clear, non-colored areas are those where the overall density at 13h00 is unchanged relative to midnight.

The image on the right depicts the major transportation transit points (the darker the location the more people it attracts, with town centers being circled) as identified by the 2013 COMTRANS study. The image shows a quasi-identical finding to the image on the left generated by our analysis of mobile big data. The green arrows running from the right-side image and the left-side image show the key points of similarity between the two.

Both images suggest how Colombo city acts a sink, pulling in people from the surrounding regions during the workday. Understanding these movements of people, and their sources, on a near-real-time basis via mobile network big data, could greatly enhance transportation planning.

5. Challenges and Future Work

[2] The 2013 COMTRANS study was done between 2012-13 in the Colombo and greater Colombo areas and comprised of data from a 40,000 household survey.

There are several challenges posed, and potential future research areas offered, by using mobile network big data for transportation planning in Sri Lanka.

5.1 Travel Motifs

One potential difficulty with taking the mobile big data approach to transportation is its reliance on generalization of human behavioral patterns to a human behavior motif holding that people "work in daytime at an office, and sleep at home at night". While this hypothesized motif was proven to be true by our data analysis, we do understand that there are segments of the working population who do not match this model. The behavioral patterns of two such segments of people would have varying impacts on our intermediate and final results, as we now outline in the next two sub-sections.

5.1.1 Individuals with Inverted Work Shifts

Individuals who stay at home in the day and go to work at night are not captured in our hypothesized motif. But it is interesting to note that this segment of the working population would not significantly alter the mathematical model due the fact that, just like the individuals who "work in daytime at an office, and sleep at home at night", these individuals who would "work at nighttime at an office, and sleep at home during the daytime" will have a night location and day location, and will commute between those two DSDs, thus not altering the overall shape of the mobility graphs. The only potential data problem would be misclassification these individuals' home locations as work locations, and vice versa.

5.1.2 Individuals with Multiple Work Locations

Individuals who have multiple work locations are not captured in our hypothesized motif. It could be proposed that these individuals – if it is assumed that they have a uniformly random-distributed number of work locations over uniformly random-distributed time periods throughout the day – would generate data points that cause a smoothing effect on the mobility graphs. One possible solution would be to identify these individuals by creating individual mobility graphs and selecting the graphs that have the most deviation from the average mobility graph (provided the stationary individuals were removed as well).

It can be argued that with these multiple-workplace individuals removed, the average mobility graph would have finer level of detail. However, it needs to be borne in mind that even without the removal of such individuals, our average mobility graph displayed the expected shape, suggesting that the proportion of individuals with multiple work locations is comparatively much smaller than the proportion of individuals with a single work location.

5.2 Using DSDs as the Smallest Geographical Unit

Given the large physical size of the DSDs, use of DSDs as the smallest geographical unit means that there can be great variation among the mobility patterns of individuals whose home DSD is the same as work their DSD. Some of these individuals will not actually be moving for their jobs, while others will have moderate movement. While this distinction may seem insignificant from a statistical perspective, it can be significant from a policy perspective, since the socioeconomic realities of a non-moving individual (e.g., a housewife, a self-employed craftsperson) will often be significantly different from those of an individual who moves a moderate distance for work (e.g., a grocery store owner). The way to solve this issue would be to work with geographical units smaller than DSDs, so as to generate finer-grained details.

5.3 Possible Extensions and Future Work

Removing individuals who do not match the "work in daytime at an office, sleep at home at night" pattern and analyzing their behavioral patterns would extend our case and provide greater understanding of the socioeconomic dynamics of mobility. And the mobility graph for the remaining "work in daytime at an office, sleep at home at night" people would be more representative of the target portion of the populace.

Moving to a smaller unit of denomination than the DSDs would also be preferable. In this respect, we gave consideration to working with Grama Niladari Divisions (GNDs), which make up the DSDs. However, there are many GNDs that do not have a BTS within their own boundary. Often a single base station provides coverage for a cluster of adjoining GNDs. Also, due to the shapes of some of the GNDs, when a GND does have a base station, that GND only contributes a small portion of the traffic going through the station.

To solve this problem, usage of voronoi cells[3] to map the base stations was considered. But then the units would not reflect any direct relationship with political or social divisions in the country. We then gave consideration to combining the two approaches, i.e., using both GNDs and voronoi cells as means to identify base stations. In this proposed solution, first a mapping function would be created between each voronoi cell coordinate and the GNDs that said voronoi cell covers. Then the voronoi cell's mobile traffic would be modeled as a decaying density function that would then be allocated to the GNDs based on area overlap.

6. Conclusions

The findings from our research suggest that in developing countries such as Sri Lanka, which have limited sensor networks, active and passive positioning data from mobile network operators (as well as real-time GPS traces from mobile phones) could revolutionize transportation management and improve the efficiency and reliability of transportation systems.

Given that the underlying data constitutes the activity of citizens' there are some natural concerns related to privacy. The analysis conducted in this paper produces only aggregate insights as the final output, albeit the process of generating those insights starts at an individual level. A discussion on modes of conducting such research whilst preserving privacy, tradeoffs, as well as on closely related (but converse) issues such as marginalization and exclusion is beyond the scope of this paper.

While we are cognizant of the fact that there are still methodological improvements required, even our preliminary data analysis has shown that mobile network big data shows promise as a source of timely and relatively cheap insights for transportation planning. Even after further refinement it is not expected that the insights derived from mobile network big data will cover all the needs of transportation planners in developing economies, nor that they will be perfect, and nor that they will remove the

---

[3] A voronoi cell in this context is the inferred spatial coverage area of a BTS assuming that the BTS is located at the center of gravity of the inferred area.

need for surveys. For example the type of MNBD utilized in this paper (i.e. CDR data) analyses will not be well suited to identifying vehicle types or traffic conditions. However public policy is about the art of the possible, not the art of the perfect. Even an incremental increase in the type and frequency of insights that are possible in resource constrained developing economies can potentially facilitate more effective planning processes.

References

Amini, L., Bouillet, E., Calabrese, F., Gasparini, L., & Verscheure, O. (2011). Challenges and results in city-scale sensing. In *2011 IEEE SENSORS Proceedings* (pp. 59–61). IEEE. doi:10.1109/ICSENS.2011.6127189

Berlingerio, M., Calabrese, F., Lorenzo, G. Di, Nair, R., Pinelli, F., & Sbodio, M. L. (2013). AllAboard: A system for exploring urban mobility and optimizing public transport using cellphone data. In *Data for Development: Net Mobi 2013*. Retrieved from http://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf

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

Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing, 10*(4), 36–44. doi:10.1109/MPRV.2011.41

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying Important Places in People's Lives from Cellular Network Data. In K. Lyons, J. Hightower, & E. M. Huang (Eds.), *Pervasive Computing: Lecture Notes in Computer Science* (Vol. 6696, pp. 133–151). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-21726-5_9

Wang, H., Calabrese, F., Di Lorenzo, G., & Ratti, C. (2010). Transportation mode inference from anonymized and aggregated mobile phone call detail records. In *13th International IEEE Conference on Intelligent Transportation Systems* (pp. 318–323). IEEE. doi:10.1109/ITSC.2010.5625188

Wu, W., Cheu, E. Y., Feng, Y., Le, D. N., Yap, G. E., & Li, X. (2013). Studying intercity travels and traffic using cellular network data. In *Data for Development: Net Mobi 2013*.

Yoo, B.-S., Chon, K., Kang, S.-P., & Kim, S.-G. (2005). Origin-destination estimation using cellular phone bs information. *Journal of the Eastern Asia Society for Transportation Studies, 6*, 2574–2588. Retrieved from http://www.easts.info/on-line/journal_06/2574.pdf