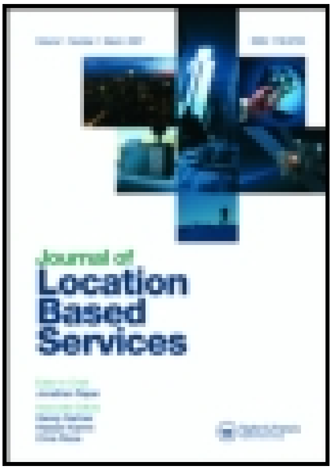

Journal of Location Based Services
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tlbs20

Investigation of travel patterns using passive cellular phone data

S. Bekhor[a] & I. Blum Shem-Tov[a]

[a] Department of Civil and Environmental Engineering, Technion - Israel Institute of Technology, Technion Campus, Haifa 32000, Israel

Published online: 14 Aug 2015.

Journal of Location Based Services, 2015
http://dx.doi.org/10.1080/17489725.2015.1066515

Taylor & Francis
Taylor & Francis Group

Investigation of travel patterns using passive cellular phone data

S. Bekhor[1] and I. Blum Shem-Tov*

Department of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Technion Campus, Haifa 32000, Israel

(*Received 20 August 2013; accepted 22 June 2015*)

This paper presents methods for inferring selected travel patterns using passive cell-phone technology. The paper handles two issues that are related to the cellular phone technology: 'zigzagging' patterns that do not represent a movement, and track recording of the closest antenna location that serves the cell-phone, which means that the information of the location of the cell-phone itself is not accurate. The home and commuter location of each user is validated at the aggregate district level. The paper shows selected results that can be used for transportation analysis, including a comparison of the results to known models from preceding years.

Keywords: travel-patterns; cellular-data; travel habits; cellular antennas

Introduction

Transportation analysis relies on survey data to provide inputs and to calibrate the mathematical equations that represent decisions people make related to their travel. A variety of passive (e.g. road sensors, cameras, floating vehicles) and active (e.g. personal or online interviews) methods are used in order to collect data. Most methods are expensive to cover a large area and suffer from reliability issues.

Literature indicates a great potential of using cell-phones as transportation sensors (Rose 2006). The cell-phone market penetration in many countries is very high. Using cell-phones as transportation sensors is very appealing because the infrastructure usually already exists and as opposed to other methods, it is not limited to small areas. Specifically in Israel, an independent telephone survey found that the cell-phone owner population is a representative sample of the entire population (Bekhor et al. 2013).

Early researches developed a methodology for extracting travel data from cell-phones (Asakura and Hato 2001) and how to use cell-phone probe data to investigate individual travel behaviour (Asakura and Hato 2004). A comparison to known systems such as floating cars and investigation of the system accuracy was performed later by Cayford and Johnson (2003). Qiu et al. (2007) reviewed studies involving simulation of data extraction. A review of practices using cell-phone as traffic probes can be found in Caceres et al. (2008). Another interesting cell-phone study is performed using billing based data to research main city movements in Estonia (Ahas et al. 2010). All studies showed a great potential using cell-phone data for transportation, however they also pointed out accuracy

*Corresponding author. Email: iblumshemtov@gmail.com

issues and the coverage limitations, which inspired the selection of travel behaviour parameters in this paper.

In the transportation field, studies have focused on the use of cell-phones as probes for estimation of aggregate travel patterns, such as travel times and speeds (Bar-Gera 2007), modal split (Wang et al., 2010), origin–destination (OD) matrices (Caceres et al. 2007; Calabrese et al. 2011) and long-distance travel patterns (Bekhor et al. 2013). Extraction of OD matrix is a classical usage of the cellular data for transportation because there is no need to find the exact route, except for the start and end points (White et al. 2004). In Israel, a national OD from cellular data was estimated by a private company (Matat 2007) and further developed in Bekhor et al. (2013).

The present paper follows this line of research and investigates aggregate travel patterns using passive cell-phone data. The paper deals with a relatively large area, which comprises the country of Israel, and also deals with a great amount of real cell-phone data, which represents a statistical sample of the whole population of Israel. In addition, this paper relies solely on cell-phone data, without making use of additional sources or other data collection techniques. The data used for the investigation includes records of about 10,000 cellular-phone users over one work week. The data source is provided by a major cell-phone company in Israel. The tracking system is based on recording events that contain a change in the position of the cell-phone with respect to a given antenna.

The extraction of the transportation information from the cell-phone data is not trivial. The massive amount of data requires working in a database environment and since the raw data collected is not designed for transportation purposes, a comprehensive data processing is needed. The data processing includes statistic data processing and sophisticated SQL queries to extract the transportation information out of the cell-phone data. Moreover, using the raw data of the cell-phone system raises some challenges. For example, the frequency of the observations is not always sufficient to infer short movements. In addition, there are two phenomena that are specifically related to the cell-phone technology: (1) network load on antenna (or other technical reasons) can cause 'zigzag' patterns where the same phone is being recorded at several antennas while actually there is no movement. (2) The phone location is recorded up to the antenna's' location so that the information of the location of the cell-phone itself, is not accurate.

Another challenge is to locate anchor points, such as place of residence and commuter locations, since the dataset does not contain any information about the cell-phone user. Therefore, several assumptions are needed to make the calculations. Those limitations are discussed in other works as well: for example, Isaacman et al. (2011) developed an algorithm that uses cellular network data to identify the important places such as home and work locations. The algorithm included predefined clustering such as work time period between 1 pm to 5 pm and home hour period between 7 pm to 7 am (Isaacman et al. 2011). After locating the home and the commuter locations, several travel movements are able to be computed, such as movement frequency (trip rates), and distribution of the movement length (trip length distribution). The results are compared to previous household survey and cell-phone survey results from the Israeli national model.

The rest of this paper is organised as follows: the next section presents the methodology devised to analyse the cell-phone data. Special attention is paid in the identification of travel movements, which is not a trivial task, given the nature of the data collected. The subsequent section presents results for main travel patterns and comparison with other surveys. The final section of the paper presents a summary of the findings and suggestions for further research.

Methodology

Data collection

During March–June 2007, a comprehensive cellular survey took place, as part of the national transportation model (Gur et al. 2009). The cellular phone positions were provided by 'Orange', one of the main three cellular phone providers in Israel. The raw data was recorded for 16 consecutive weeks between March 7 and July 2, 2007. Every week, an average sample of 10,200 cellular phone numbers was randomly drawn from one cellular phone provider. The private company that performed the project for the Ministry of Transport (Matat 2007) provided one week of the raw data from the survey, which was used as the basis for this paper. The raw data included the following information: (1) sensor ID (which is different from the phone number, to avoid personal identification), (2) the cellular antenna location that is serving the cell-phone during the record time, and (3) the record time. Every row in the database represents a sample of a sensor ID in the network at a specific time. In the case of no movement and no call, the minimum sample frequency is once every two hours, otherwise, it is every antenna handover (which might indicate a movement). During calls, the sample frequency increases and many identical measurements of antenna locations are observed.

The minimum recording frequency was set to 2 h. This value is related to cellular phone provider policy. This means that in case that no cellular phone movement is observed (and the phone is on), the company scans the phone and records its position every 2 h. In case the cellular phone changes antenna, the movement is recorded. In addition, when a person speaks in the cellular phone, the recording frequency is very high.

Antenna spatial resolution

The phone location is recorded up to the cell-id or antenna's' location so that the information of the location of the cell-phone itself is not accurate. The distribution of the antenna locations in Israel is not homogenous, because populated areas have more antennas than isolated areas, but it covers all the country. This limitation may affect the accuracy of movements in rural areas, in which short trips may be captured by the same antenna, but the majority of trips within the whole population of Israel will not be affected because of multiplicity of other types of antennas that are being used more often those days. Another limitation that comes from the resolution of the antennas is identifying slow modes such as walking, biking, and so on. Since the resolution is not sufficient to determine the mode we did not deal with this subject in the paper.

An independent survey conducted by a private company (Matat 2007) shows that the penetration rate of cellular phones in Israel is sufficient to represent the whole population. Prior to the data collection, a random CATI survey was conducted to check the penetration rate by different socio-economic characteristics. The results of this survey were reported in a previous paper (Bekhor et al. 2013) and are briefly summarised here: 83% of aged 8 and up are equipped with at least one cellular phone, and within the ages 18–60 this rate increase to 92.6%. There is a cell-phone to 95% of the employees, 93% of the students, 99% of the working students and 100% of the soldiers.

Data preparations

The extraction of the transportation information from the cell-phone data is not trivial. The massive amount of data (more than 4 million records for 10,000 phones over one week) requires working in a database environment. The preliminary analysis includes identification of incorrect observations such as those without geographical information or erroneous coordinate measurements (records with missing latitude/ longitude values). All the proper locations are projected to the new Israeli grid (from spheroid coordinate system) and some random sensor observations are visually tested in ArcMap.

After the preliminary analysis, a filtering process is applied to remove redundant data which is sourced in the 'Zigzagging' patterns phenomena. The 'Zigzagging' patterns happen when the phone is being recorded at several antennas sequentially because of network load in one antenna or other technical reasons. Since the 'Zigzagging' patterns are created by repeated antenna handovers, they are assumed to happen when the cell-phone is located in the congruent coverage area of two antennas. Although the 'Zigzagging' is recorded as a handover, there is no physical movement of the cell-phone, and, therefore, the data is not relevant to the research. Circular 'Zigzagging' patterns, where the first and last antennas are the same, are detected automatically and removed from the database. The max time difference that is used to define an unwanted circular 'Zigzagging' pattern is calculated from the time difference between two successive records for the same sensor ID in the database and is set to one minute (Figure 1).

The 'Zigzagging' pattern is presented in Table 1 and Figure 2, for a single sensor ID 1057069001. Figure 2 shows the actual antenna locations and the distance between them. The database records show three different antenna locations (Table 1). In a very short period of time, the network antenna switches between 1881, 782 and 39091 as indicated by the short-time difference and consequently high speed (unusual for a dense urban area with local streets). Note that the last record in Table 1 shows that the Sensor ID is 'back' to the original position, so in this case the zigzagging pattern does not allow inferring if there is an actual movement or the sensor ID was in the same physical location.

In order to properly map the records that represent actual transportation movements, another filtering process is applied. The filtering process is done according to the maximum time difference between every two sequential samples ($\Delta T_{\max}$). In order to calculate $\Delta T_{\max}$, a statistical approach is used. A histogram of the time differences between every two sequential measurement, ΔT_i, is calculated (Figure 3), and after analysis, it is found that more than 80% of ΔT_i were less than 100 s. Accordingly, $\Delta T_{\max}$ is set to be 100 s (the greater the time difference between two sequential samples the less accurate the results will be. A time difference which is greater than 5 min could lead to errors in movements calculation, so if we take 95% of all the ΔT_i up to 300 s, we obtain 100 s). In addition, the histogram clearly shows the constant 2-h system value of time between measurements when there is no handover.

Data analysis

The removal of samples that are not relevant for transportation travels, allows the beginning of travel patterns extraction from the database. However, working in a large database environment requires the usage of efficient operations to allow the algorithms to execute in reasonable time. Since the raw data includes more than 4 million records, efficient SQL operations and queries are used for the data processing. The analysis is done in two phases: First, all results are calculated on small datasets, and then, after getting results, they are calculated again on the whole dataset and compared.

The calculation of the sensor ID movements needs to be preceded with several SQL operations on the database. First, for each sensor ID, all the data records are arranged by

Figure 1. Antenna distribution in Israel.

Table 1. Zigzagging example in the database.

Sensor ID	Date	Time	Antenna ID	Elapsed time (min)	Speed (km/h)
1057069001	25 March 2007	15:53:26	1881		
1057069001	25 March 2007	15:54:46	782	1.33	133.7
1057069001	25 March 2007	15:55:40	39091	0.90	118.7
1057069001	25 March 2007	15:56:38	782	0.97	110.5
1057069001	25 March 2007	16:05:41	1881	9.05	19.7

the timestamp. Then, for each sensor ID, every two sequential records are joined to one data row in order to make difference calculations of distance, time, etc. Finally, movements extraction algorithm is applied so that movements with the home and commuter locations are inferred for each sensor ID.

Home and commuter locations

Since the dataset does not contain any information about the cell-phone user, several assumptions are made to locate anchor points such as place of residence and commuter location. The cell id that represents the home location of each sensor ID is inferred assuming that each person stays at home during the night. For each sensor ID, a constant number of records just before a certain time and just after that time are excluded from the database for the whole surveillance period (the times tested where between 9 pm and 3 am). For each sensor, the amount of records within this time period is being summarised and set to be the home location of each sensor. This calculation was done repeatedly for a few times (for each tested hour). The midnight time was found to be the best reference

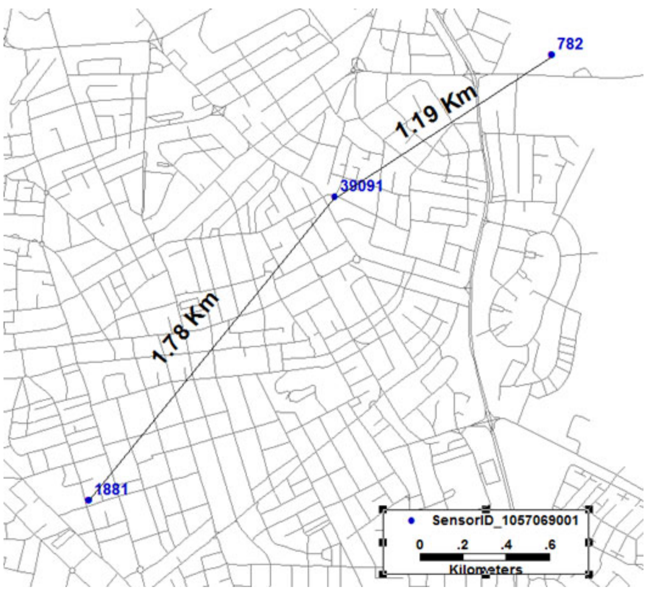

Figure 2. Data problem – antenna locations.

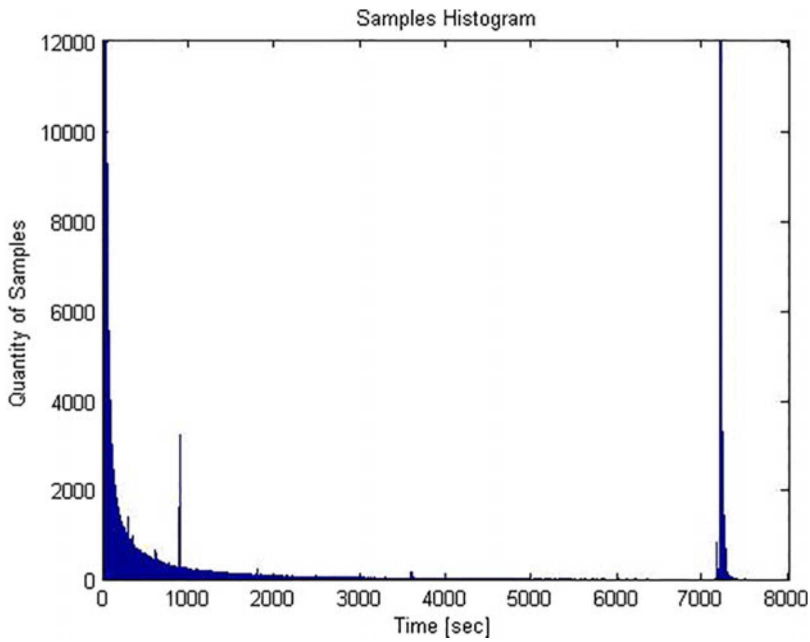

Figure 3. Time difference histogram.

time, as presented later in the results section. The reference time is the time in which the cell phone was at home.

Another approach to finding the residence was to calculate the home location by frequency of observations (the place with the highest number of observations was considered as the home location). This method failed because 20% of the sensors got the same cell as a representative of both home and commuter locations. Note that the calculation of the home and commuter location is precise up to the antennas' sector. Each antenna has three sectors, each responsible to 120° so there is a total of 360° of coverage area around a certain antenna. Given this relatively good precision, it is not acceptable that about one-fifth of the population has the same sector for both home and commuter location. As a consequence, the home location was calculated using the time reference only.

The commuter location (e.g. workplace or education place) is calculated in two separate ways. In the first calculation, the second most recorded location is stored (hence calculated by frequency), while the second calculation is performed similarly to the home location calculation, only assuming that each sensor ID is at the commuter location at 11 am (hence commuter calculated by time). Accordingly, two movement logs are extracted from the cell-phone data. Both logs are compared, and in cases that the locations do not match, the most frequent observation out of the two is set as the commuter location.

Movements

Movements are calculated after home and commuter are inferred. In order to calculate the movements, an additional index is added to the database (*Stay* column in Table 2). *Stay* is

Table 2. Example of processed database.

Sensor ID	Location coordinates	Next location coordinates	Time difference (s)	Stay
1234	(X_1, Y_1)	(X_1, Y_1)	30	1
1234	(X_1, Y_1)	(X_2, Y_2)	10	0
1234	(X_2, Y_2)	(X_3, Y_3)	40	0
1234	(X_3, Y_3)	(X_4, Y_4)	20	0
1234	(X_4, Y_4)	(X_4, Y_4)	35	1
1234	(X_4, Y_4)	(X_4, Y_4)	40	1

an indicator for the change of coordinates, when *Stay* is zero there is a change of location between two sequential measurements. This indicator helps calculation movements that might represent movements.

The algorithm for calculating suspected movement is based on the following rules:

- Beginning of movement occurs when the first *Stay* index position in the database is equal to one, following a record with movement (*Stay* is zero).
- The duration of a movement (or suspected trip) is set for consecutive records with *Stay* equals zero.
- A stop is assumed for records with *Stay* equals one for less than 60 s.
- Ending of movement occurs if there is no change of location during two sequential records or a stop for more than 60 s.

According to the above definitions, in this example the movement starts at the second row and finishes at the fourth row. The departure time was taken as the time of the first record of the trip, and the duration was the time difference between the last record associated with the trip and the first one.

Figure 4 shows an example of the recorded events for a single sensor ID. The suspected home and commuter locations are circled. The labels shown are the amount of samples per antenna. The home location (the upper circle) was sampled 128 times, and the commuter location (the lower circle) was sampled 91 times.

Results

This section presents selected results of main transportation characteristics that were extracted from the data. The analysis of the results focuses on general transportation parameters that represent the transportation habits of the whole Israeli population. Whenever possible, the results are compared to other existing national surveys, as follows:

- National Census, 2008 (home and work locations).
- National Travel Habits Survey 1996/1997 that was conducted by the Central Bureau of Statistics (CBS).
- A national model to predict passengers' trips based on cell-phones location changes (Matat 2007).

Home and commuter locations

As mentioned in the methodology section, the process of calculating the home location was performed several times, each time with a different reference hour for the home

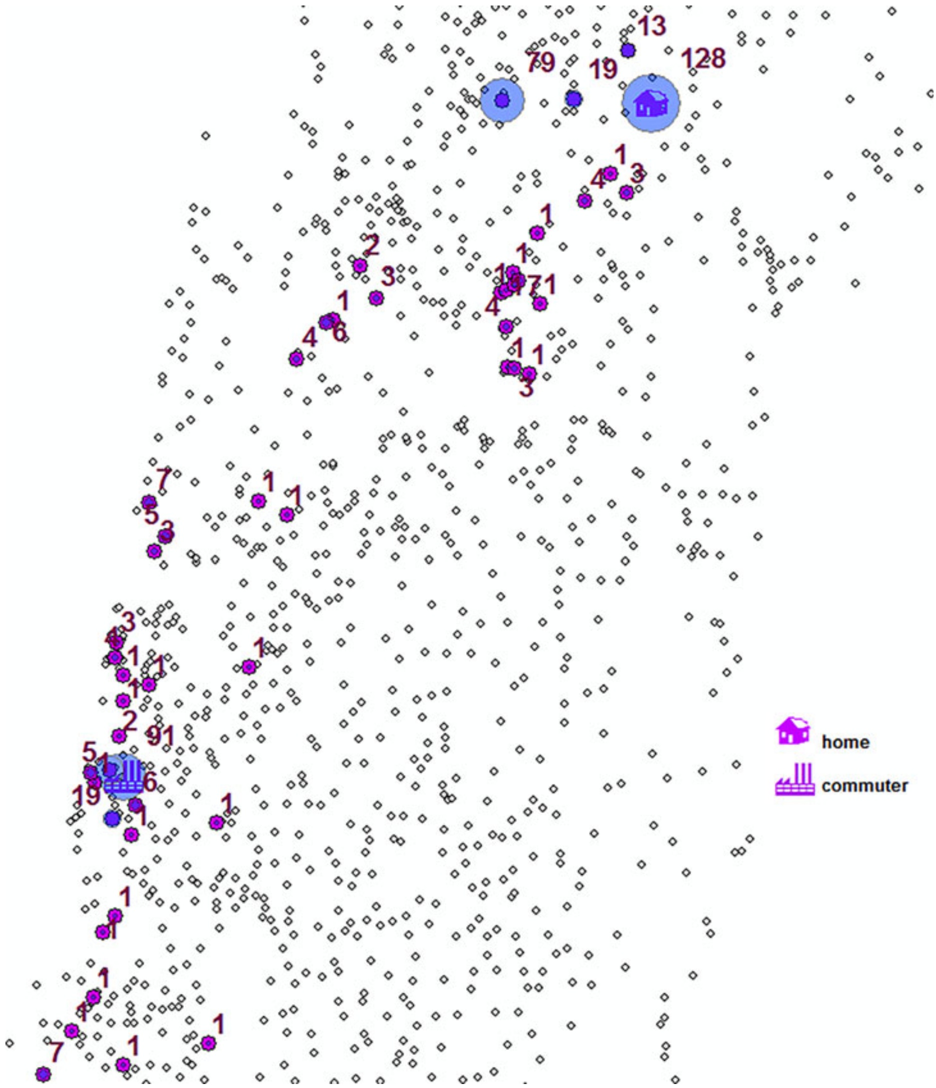

Figure 4. Example of the movements for a single sensor ID.

location as follows: 9:00 pm, 10:00 pm, 11:00 pm, midnight; 1:00 am, 2:00 am, and 3:00 am. The results show an overall good correlation with the population data from the 1998 Census. Although there is no significant difference between the specific hours assumed for the home location, the best correlation found was at midnight.

Table 3 compares the distribution of home location for each hour and the population according to the 1998 Census. The Pearson's correlation between the population distribution and the home locations shows high values (between 0.93 and 0.94). This data is also illustrated in Figure 5.

First, it can be seen that the reference time at night has no actual influence on the distribution of the calculated home locations. Second, the distribution of the calculated

Table 3. Distribution of the calculated home locations and the actual population according to area.

District	Sub district	3:00 (am)	2:00 (am)	1:00 (am)	0:00 (am)	11:00 (pm)	10:00 (pm)	9:00 (pm)	Population
Jerusalem (%)	–	10.6	10.5	10.6	10.6	10.7	10.8	10.7	12.2
North (%)	Zefat	1.1	1.1	1.0	1.0	1.1	1.1	1.1	1.4
	Kinneret	0.9	1.0	0.9	1.0	0.9	0.9	0.9	1.4
	Izrael	4.3	4.3	4.2	4.5	4.3	4.2	4.2	6.0
	Ako	4.5	4.4	4.4	4.8	4.4	4.5	4.5	7.5
	Golan	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5
Haifa (%)	Haifa	8.0	7.9	8.0	8.2	8.0	8.0	7.9	7.2
	Hadera	2.8	2.8	2.8	2.9	2.9	2.9	2.9	4.8
Centre (%)	Sharon	4.4	4.4	4.5	4.3	4.7	4.7	4.6	5.3
	Petach-Tikva	9.2	9.3	9.4	9.3	8.9	8.9	9.1	8.1
	Ramle	4.4	4.3	4.5	4.5	4.4	4.4	4.3	3.8
	Rehovot	6.4	6.5	6.4	6.3	6.3	6.2	6.3	6.7
Tel Aviv (%)	–	23.0	23.0	23.3	21.8	22.3	22.4	22.5	17.0
South (%)	Ashkelon	5.5	5.5	5.3	5.7	5.7	5.6	5.7	6.3
	Bear Sheva	7.8	7.8	7.7	8.0	7.9	8.0	7.9	8.1
Yosh (%)	–	6.7	6.7	6.6	6.8	7.0	6.9	6.9	3.8
Total (%)		100	100	100	100	100	100	100	100
Pearson's correlation		0.932	0.931	0.930	0.940	0.934	0.936	0.935	

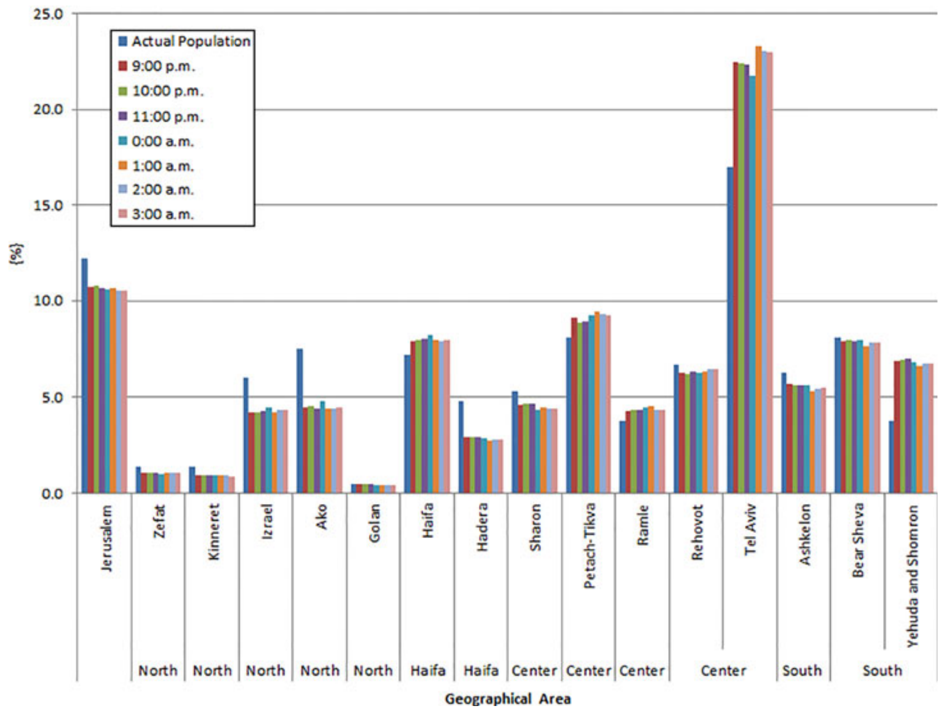

Figure 5. Distribution of the calculated home locations and the actual population according to area.

home locations resembles to the distribution of the 1998 Israeli Census. The slight differences can be explained by the difference in the percentage of cellular phone users of the 'Orange' provider over the different areas.

Table 4 shows the percentage of the cellular phone users in the research survey (Matat 2007). The second column is the percentage of the 'Orange' provider in the sample (around one-third of the market), and the third column shows the high penetration of the

Table 4. Percentage of cellular phone users (Matat 2007).

Area	'Orange' users among age 8 + (%)	Population with cellular phone among age 8 + (%)
Northen Galille	35.0	89.4
Haifa	23.5	85.3
Lower Galille	20.8	86.1
Central district-north	32.4	87.5
Central district-south	31.0	89.0
Tel Aviv	29.8	87.3
Yosh	30.4	91.3
Jerusalem	19.9	81.3
South Israel	19.3	81.3
Bear Sheba	25.1	86.0

cellular phone users overall. It can be seen that in the areas that the correlation between the calculated home location and the population percentage is varying, the percentage of the cellular phone users also varies. For example, in the northern district fewer home locations are calculated from the population data, but according to Table 4, there are also fewer 'Orange' cellular users in the Haifa and Lower Galillee sub-districts. On the other hand, in the Tel-Aviv district, more home locations are calculated via the cellular data from the population data, which can also be explained by the higher percentage of cellular users in the Tel-Aviv area.

Table 5 and Figure 6 present the distribution of the calculated commuters with the distribution of the work from the 1998 Census. The Pearson's correlation value is quite high (0.97) for work location.

The histogram and table show the distribution of the population (residence) from the 1998 Census and the work locations compared to the calculated commuter location. They show that the distribution of calculated commuters and residence places at the district level are quite similar (Pearson's correlation 0.94). This is explained by the fact that most residents work and live at the same district, and the commuter in many cases is the work place.

Movement parameters

Table 6 presents the average amount of movements ('trips') per person, per day with their lengths, durations, and average speed. Matat survey shows an increase of 38% in long trips (over 25 km) compared to the Israel CBS household survey from 1998. On the other hand, Matat survey did not consider short distance (less than 3 km) trips, and for this reason the average number of trips is similar to the CBS survey. Note also that the Israel CBS household survey was conducted in 1998, around 10 years before the sampling of the data that is used here (2007). According to that, it can be assumed that the parameters show consistency with all the other models that were reviewed.

Table 5. Distribution of the commuter and work locations and the population.

District	Sub district	Commuter 2008 census	Workplace cellular survey
Jerusalem (%)	–	10.6	9.4
North (%)	Zefat	1.0	1.4
	Kinneret	1.2	1.2
	Izrael	4.3	5.0
	Ako	4.8	6.0
	Golan	0.4	0.4
Haifa (%)	Haifa	8.5	8.0
	Hadera	2.7	4.4
Centre (%)	Sharon	4.2	5.4
	Petach-Tikva	8.9	9.5
	Ramle	4.5	4.1
	Rehovot	6.6	8.1
Tel Aviv (%)	–	22.2	20.7
South (%)	Ashkelon	5.6	6.4
	Bear Sheva	7.7	6.8
Yosh (%)	–	6.5	3.3
Pearson's corellation with commuter		1.00	0.97

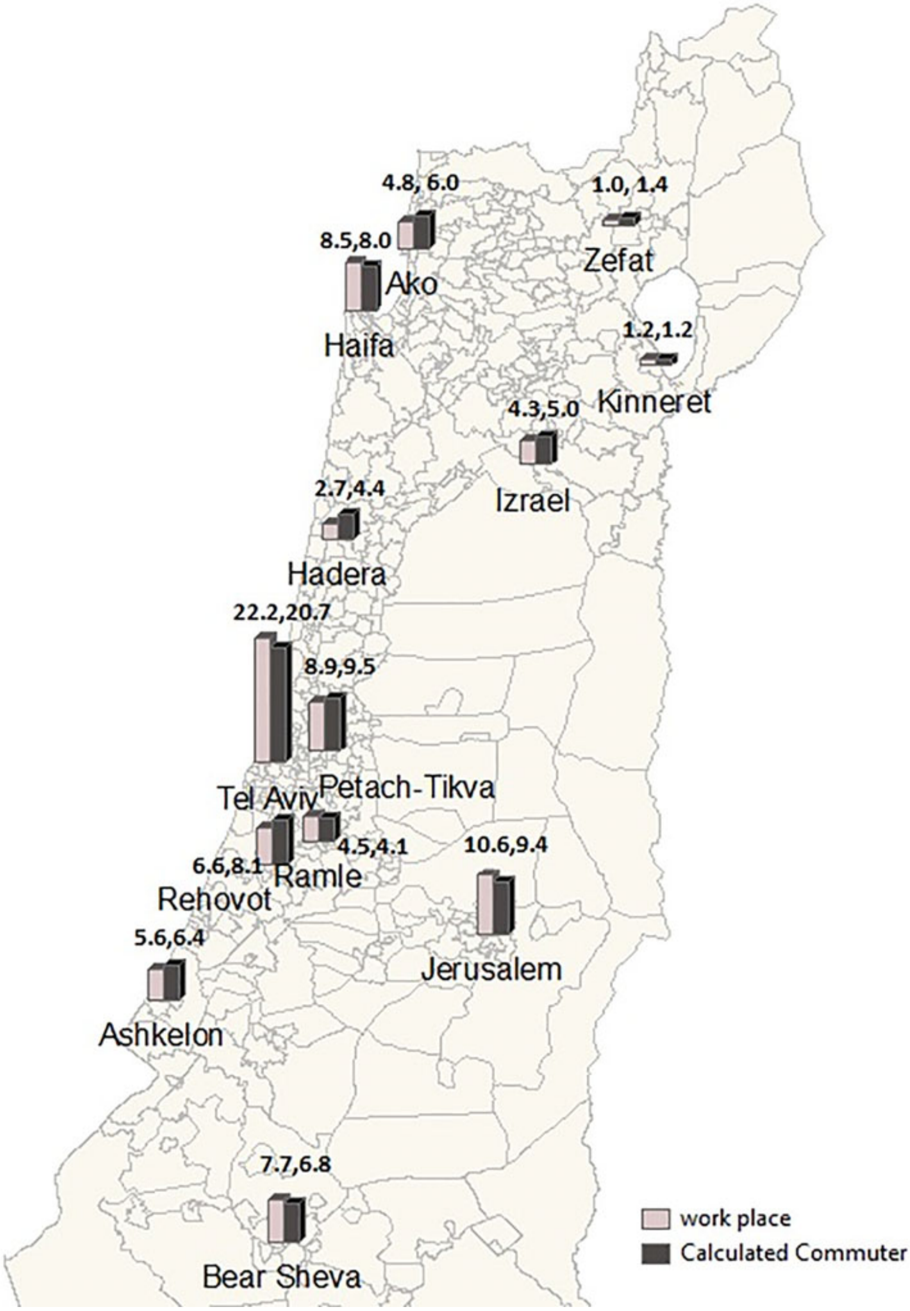

Figure 6. Distribution of the calculated commuter and work locations.

Sample analysis

Figure 7 presents the distribution of the weekly average of number of samples in 24 h (In the Matat survey, the average amount of records per day is displayed for the whole

Table 6. Average of trip/movement properties per person, per day.

Survey		Number of trips	Trip distance (km)	Trip duration (min)	Trip average speed (km/h)
Matat survey (2007)		~2.0	–	–	–
Israel CBS household survey (1998)		2.0	12.6	26.7	28.3
This Research (data from 2007)	Commuter calculated by *frequency*	2.3	14.1	23.6	35.8
	Commuter calculated by *daytime*	2.4	13.6	22.6	36.1

16 weeks database). Throughout the whole research all the calculations are made for two movements/trips logs: one, in which the commuter was calculated by time, and another in which the commuter was calculated by frequency. According to this logic the figure shows both results in comparison to the results of the Matat survey. The dotted line represents the distribution from the Matat survey and the other dashed lines are the results from this research. It is also possible to observe a peak in the '11–20' category, which can be explained by the 2-hour time lag effect of the consecutive observations without movement. The meaning of that observation is that about 20% of the records are not relevant for movement measurements (of course the '0–10' category is not relevant as well but the pick at '11–20' represents the cell phones that did not move and were recorded every 2 hours). However, the rest 80% of the records are relevant and quantitatively sufficient for the research.

Moreover, since both this research (1 week) and Matat's survey (16 weeks) results correlates, it might indicates that 1 week of measurement can be sufficient for representing the whole 16-weeks survey.

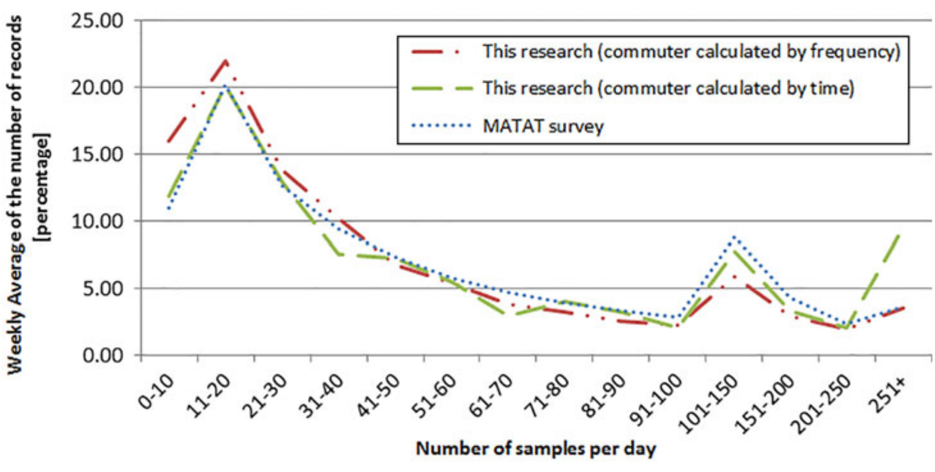

Figure 7. Distribution of the weekly average of records in 24 h.

Movements length distribution

Figure 8 presents the percentage of movements by distance that were calculated in this research compared to the Matat survey (dotted line). It can be seen that this research results are consistent with Matat results for trips longer than 50 km, and that for trips that are less than 50 km, there is a slight difference between them. The difference can be explained by the fact that Matat research did not considered short-distance trips (less than 5 km) and metropolis trips, while this research considered all trips/movements, both short and long.

Figure 9 presents the average trip/movement length by the time period as was calculated in this research (dashed lines) compared to the Matat survey (dotted line).

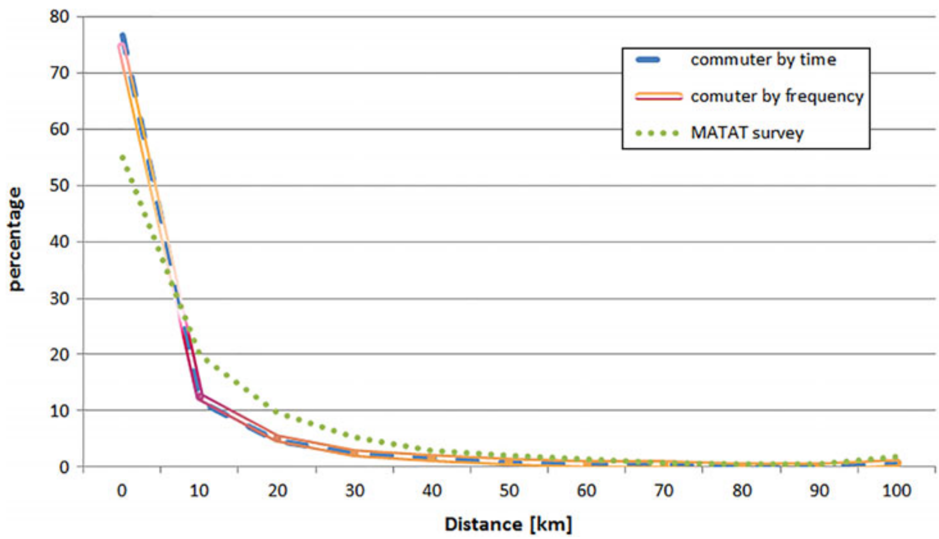

Figure 8. Trips/movements by distance.

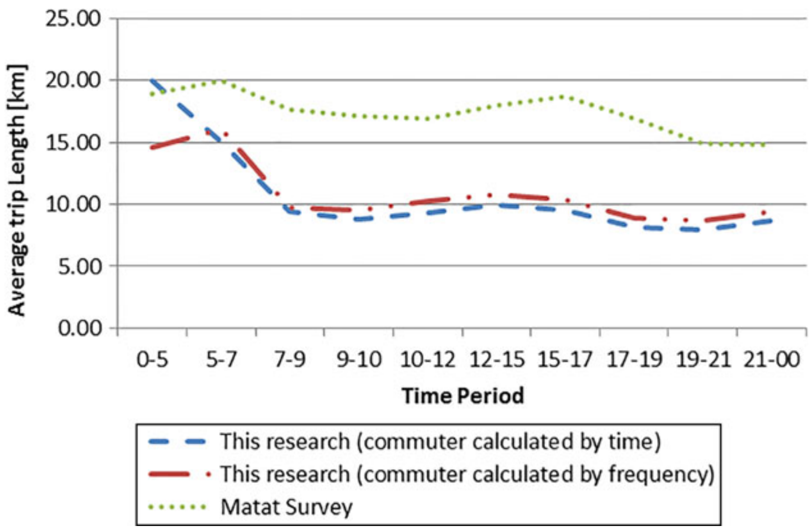

Figure 9. Average trip/movement length by time period.

The association of trips with periods was according to the departure time. The line trends are consistent but there is an offset in the *y* axis. That offset can be explained by the same argument as in Figure 8. The results reflect the common trend of longer movements in the rush hours (6–10, 15–19).

Trips/movements distribution by departure time

Figure 10 presents the trips/movements distribution by the departure time. Here, the trips/movements are not categorised and all trips/movements are considered. Except for the morning hours, the general trends of this research results lines and the Israel CBS lines are alike. The difference in the morning hours is explained because of a lack of calls during the morning hours. The assumption is that people do not tend to make social calls in the morning hours and that offices are closed so people do not make arrangements calls during those hours. The other differences may be caused by the time difference between the measurements or changes in travel habits at certain time periods over the years.

Trips/movements distribution by time for the main destination

In this section two main destinations are analysed. The first is the commuter (e.g. worker, student) destination which is assumed to be travelling in the morning hours

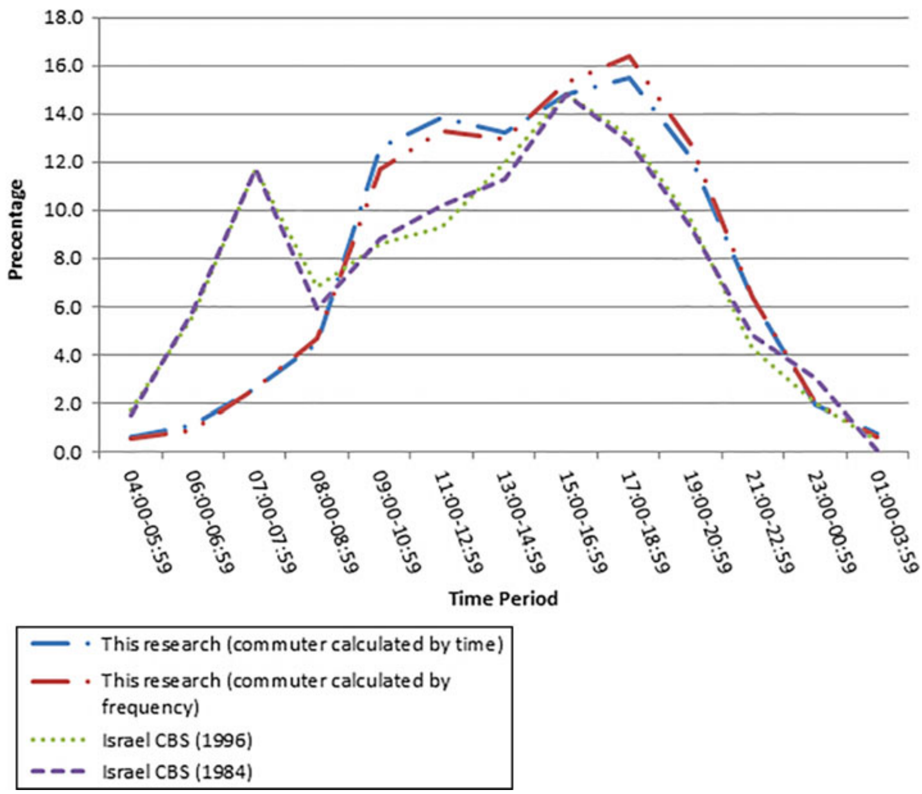

Figure 10. Trips/movements distribution by departure time.

Table 7. Trips/movements to the commuter destination.

Commuter calculated by *frequency*			Commuter calculated by *time*		
Time	Amount	Percentage	Time	Amount	Percentage
01:00–4:59	38	0.91	01:00–4:59	77	1.20
05:00–6:59	108	2.57	05:00–6:59	188	2.94
07:00–08:59	417	9.95	07:00–08:59	579	9.04
09:00–10:59	532	12.69	09:00–10:59	785	12.26
11:00–12:59	559	13.34	11:00–12:59	874	13.65
13:00–14:59	613	14.63	13:00–14:59	948	14.81
15:00–16:59	639	15.25	15:00–16:59	908	14.18
17:00–18:59	498	11.88	17:00–18:59	813	12.70
19:00–20:59	407	9.71	19:00–20:59	655	10.23
21:00–22:59	282	6.73	21:00–22:59	415	6.48
23:00–00:59	97	2.31	23:00–00:59	160	2.50
Total	4190	100	Total	6402	100
Sum between 05:00 and 08:59: 12.52%			Sum between 05:00 and 08:59: 11.98%		

(06:00–09:00), and the second is the home destination, which is assumed to be travelling to during the afternoon-evening time (13:00–21:00). Table 7 presents the quantity of movements to the commuter destination by time calculation and by frequency calculation. The Israel CBS survey states that about 66.6% of trips to work happen between 06:00 and 09:00 (about 4 million trips per day). This research is not consistent with that result: and the numbers calculated are significantly smaller (12.52% by frequency and 11.98% by time), that is, about 838,600 trips/movements per day.

Two explanations are proposed for that inconsistency: (1) – In this research the commuter indicates the most frequent activity (after home), that could be not only work, and (2) – There might be a lack of calls during the morning hours as was seen before.

Table 8. Trips/movements to the home destination.

Commuter calculated by *frequency*			Commuter calculated by *time*		
Time	Amount	Percentage	Time	Amount	Percentage
01:00–4:59	67	1.37	01:00–4:59	117	1.48
05:00–6:59	48	0.98	05:00–6:59	128	1.62
07:00–08:59	136	2.78	07:00–08:59	257	3.24
09:00–10:59	308	6.30	09:00–10:59	596	7.52
11:00–12:59	421	8.61	11:00–12:59	719	9.07
13:00–14:59	561	11.48	13:00–14:59	924	11.66
15:00–16:59	724	14.81	15:00–16:59	1114	14.06
17:00–18:59	929	19.01	17:00–18:59	1459	18.41
19:00–20:59	884	18.09	19:00–20:59	1298	16.38
21:00–22:59	601	12.30	21:00–22:59	980	12.37
23:00–00:59	208	4.26	23:00–00:59	332	4.19
Total	4887	100	Total	7924	100
Sum between 13 and 21: 63.39%			Sum between 13 and 21: 60.51%		

Table 8 presents the quantity of trips/movements to the home location by time calculation and by frequency calculation. Here, there is a clear correlation with the Israel CBS survey. The Israel CBS survey found that 70% of the trips home happens between 13:00 and 21:00 while this research data indicates a correlation by the observation of over 60% in both frequency calculation and time calculation.

Conclusion

This paper shows that it is possible to complement standard transportation data sources that are expensive to cover large areas with high reliability. It also shows that there is a feasible way to extract transportation data from cell-phone data. The advantages of this research inferred from the wide coverage area, the independency in other sources and the given conclusion that the widespread use of cell-phones is suitable to represent the whole population. Along with that, there are some limitations using passive cell-phone data for transportation. The resolution depends on the antennas' location but can be solved by obtaining the signal strength and angle from the cell phone provider. The price of relatively cheap data collecting is costly in the data processing and the need for several assumptions to make the data ready for transportation applications. Therefore, the travel habits chosen to be analysed in this paper were due to the data limitations.

Most of to the results correlate with the surveys that they are compared to. There is only a significant difference found: a lack of movements during the morning hours. This digression might be explained by the tendency to make fewer calls during the early morning period, since most offices are closed and it is not customary to make calls during the early morning hours. This bias can be handled by using active sampling during those hours or by using passive data with more frequent measurements (the frequency of the measurements is determined by the cellular phone provider). It is also noticeable that the results of the paper are slightly different from those of Matat, because the paper considered all movements and is not limited only to long-distance trips as in Bekhor et al. (2013).

Future work and potential use

There are many more research possibilities with cell-phone data for the purpose of transportation information extraction. Using passive cellular data with additional information such as the signal strength, and using more kinds of antennas can help in extracting travel routes, detect public transportation (train and bus drives) and more accurate trips/movements parameters. Moreover, with the massive penetration of third generation cell-phones to the market there is an advantage to integrate passive information that is already collected by the cell-phone provider with more probes that are already found in the cell phone such as GPS, Wi-Fi and more options of collaborating with the users. Such collaboration can be driven by an application or designated software allowing users to specify their age, sex, home location, etc. Integrating GIS data about national parks, land-use and other geospatial information can obtain more travel behaviour patterns and can also be used for national security and emergency planning. Those results can be implemented in almost any country with cell-phone infrastructure, depending on the cell-phone penetration among the population.

Acknowledgements

We thank three anonymous reviewers for the several constructive comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research has been supported in part by [grant number 1479/13] from the Israel Science Foundation administered by the Israel Academy of Sciences.

Note

1. Email: sbekhor@technion.ac.il

References

Ahas, R., A. Aasa, S. Silm, and M. Tiru. 2010. "Daily Rhythms of Suburban Commuters' Movements in the Tallinn Metropolitan Area: Case Study with Mobile Positioning Data." *Transportation Research Part C* 18 (1): 45–54. doi:10.1016/j.trc.2009.04.011.

Asakura, Y., and E. Hato. 2001. "Behavioral Monitoring of Public Transport Users Through a Mobile Communication System." *Journal of Advanced Transportation* 35 (3): 289–304. doi:10.1002/atr.5670350307.

Asakura, Y., and E. Hato. 2004. "Tracking Survey for Individual Travel Behavior Using Mobile Communication Instruments." *Transportation Research Part C* 12 (3–4): 273–291. doi:10.1016/j.trc.2004.07.010.

Bar-Gera, H. 2007. "Evaluation of a Cellular Phone-Based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel." *Transportation Research Part C* 15 (6): 380–391. doi:10.1016/j.trc.2007.06.003.

Bekhor, S., Y. Cohen, and C. Solomon. 2013. "Evaluating Long-Distance Travel Patterns in Israel by Tracking Cellular Phone Positions." *Journal of Advanced Transportation* 47 (4): 435–446. doi:10.1002/atr.170.

Caceres, N., J. P. Wideberg, and F. G. Benitez. 2007. "Deriving Origin–Destination Data from a Mobile Phone Network." *IET Intelligent Transportation Systems* 1 (1): 15–26. doi:10.1049/iet-its:20060020.

Caceres, N., J. P. Wideberg, and F. G. Benitez. 2008. "Review of Traffic Data Estimations Extracted from Cellular Networks." *IET Intelligent Transport Systems* 2 (3): 179–192. doi:10.1049/iet-its:20080003.

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

Cayford, R., and T. Johnson. 2003. ".Operational Parameters Affecting the Use of Anonymous Cell Phone Tracking for Generating Traffic Information." Proceedings of the 82th Annual Meeting of the Transportation Research Board, Washington, DC.

Central Bureau of Statistics. 1998. "*National Travel Habits Survey 1996/7*, (Hebrew)." Report prepared for the Israeli Ministry of Transport.

Gur, Y., S. Bekhor, C. Solomon, and L. Kheifits. 2009. "Intercity Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone Data." *Transportation Research Record: Journal of the Transportation Research Board* 2121: 145–151.

Isaacman, S., R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. 2011. "Identifying Important Places in People's Lives from Cellular Network Data." In *Pervasive Computing*, 133–151. Springer Berlin Heidelberg.

Matat. 2007. "A national cell-phone based model for travel forecasting (in Hebrew)." Intermediate Report to the Israeli Ministry of Transport.

Qiu, Z., J. Jin, P. Cheng, and B. Ran. 2007. "State of the Art and Practice: Cellular Probe Technology Applied in Advanced Traveller Information System." Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Rose, G. 2006. "Mobile Phones as Traffic Probes: Practices." *Prospects and Issues. Transport Reviews* 26 (3): 275–291.

Wang, H., F. Calabrese, G. Di Lorenzo and C. Ratti. 2010. "Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records." Proceedings of the 13th IEEE Conference on Intelligent Transportation Systems, pp. 318–323.

White, J., J. Quick, and P. Philippou. 2004. "The Use of Mobile Phone Location Data for Traffic Information." Proceedings of the 12th IEE Road Transport Information and Control, pp. 321–325.